AFRL-RH-WP-TR-2009-0049

Predictors of Interpersonal Trust in Virtual Distributed Teams

Younho Seong
North Carolina Agricultural & Technical State University
Department of Industrial & Systems Engineering
1601 E. Market Street
Greensboro NC 27411

September 2008

Final Report for June 2006 to September 2008

Air Force Research Laboratory
711th Human Performance Wing
Human Effectiveness Directorate
Warfighter Readiness Research Division
Logistics Readiness Branch
Wright-Patterson AFB OH 45433-7604

NOTICE AND SIGNATURE PAGE

AFRL-RH-WP-TR-2009-0049 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNED//
PAUL D. FAAS
Work Unit Manager
Logistics Readiness Branch

//SIGNED//
DANIEL R. WALKER, Colonel, USAF
Chief, Warfighter Readiness Research Division
Human Effectiveness Directorate
711th Human Performance Wing
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
September 2008	Final	June 2006 – September 2008

4. TITLE AND SUBTITLE Predictors of Interpersonal Trust in Virtual Distributed Teams	**5a. CONTRACT NUMBER** FA8650-06-1-6743 **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER** 62202F
6. AUTHOR(S) Younho Seong	**5d. PROJECT NUMBER** 7184 **5e. TASK NUMBER** D2 **5f. WORK UNIT NUMBER** 7184D205
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) North Carolina Agricultural & Technical State University Department of Industrial & Systems Technology 1601 E. Market Street Greensboro NC 27411	**8. PERFORMING ORGANIZATION REPORT**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Materiel Command Air Force Research Laboratory 711th Human Performance Wing Human Effectiveness Directorate Warfighter Readiness Research Division Logistics Readiness Branch Wright-Patterson AFB OH 45433-7604	**10. SPONSOR/MONITOR'S ACRONYM(S)** 711 HPW/RHAL **11. SPONSOR/MONITOR'S REPORT NUMBER(S)** AFRL-RH-WP-TR-2009-0049

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
88th ABW/PA cleared on 17 June 2009, 88ABW-2009-2636

14. ABSTRACT
Human trust in virtual distributed teams plays a critical role in performing missions/tasks given the limited resources. This study attempted to review the available literature with regards to the human interpersonal trust characteristics to identify important traits or factors when temporary mission oriented teams must be assembled. Based on the review of the theories, a theory of human trust, trust continuum/grid/cube, was developed and discussed this study. Finally, an experimental framework to test the proposed theory was developed, however the experiment was not conducted due to the time limitation.

15. SUBJECT TERMS Interpersonal Trust, Virtual Teams, Distributed Teams, Theory of Trust

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Paul D. Faas
U	U	U	SAR	130	**19b. TELEPHONE NUMBER** *(include area code)* NA

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18

THIS PAGE LEFT INTENTIONALLY BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

To support human judgment and decision-making performance in dynamic, uncertain environments, many technological advances have been implemented. A technology that has been widely applied in the real-world is automation. Automation to assist decision-making has been developed in various forms, ranging from simple mechanical devices that display the status of a subsystem in order to assist humans in diagnosing the entire system, to very complex forms of automation that may develop their own decision-making strategy, such as using the artificial intelligence techniques. In a domain where the environmental states are not directly observable, various forms of intermediaries have been designed to acquire necessary information about the environment and to assist human judges in decision-making. It is the job of these automated aids to collect necessary information and process it to generate more useful information that can be considered by human operators. Note that even with a high level of reliability and accuracy in an automated aid, the presence of uncertainty in the environment, which may be transferred into the diagnostic estimates of the automated aids, causes difficulties in judgment tasks. Therefore, it is necessary to design automation that can collect or measure necessary information reliably, and consistently, so that it can positively affect human judgment performance.

In situations where the courses of actions are irreversible or lead to high risk, or the level of uncertainty is very high, the final authority for judgments tends to be reserved for human operators. Thus, automated systems in these cases are intended to assist human operators by providing informational advice about the states or events of the environment. Then, human operators can either veto or approve the advice provided by the automated systems. That is, human operators must first estimate the states of the environment and then have discretion whether to accept or reject the estimates from the automated systems. In this research, we hypothesized that this utilization of automated systems depends, in part, on the human operators' trust in the systems. As the automated system provides “good” estimates of the environment and demonstrates “good” behavior, one would expect human operators to trust in the automated system more, and consequently rely on the system more.

The study of trust has a long history; however, that history is not rich in empirical studies. Deutsch (1958) made an attempt to define and examine potential characteristics of the term from the sociological perspective. Other studies since then have been performed to understand the role of trust in interpersonal relationships. Recent studies in a sociological context define trust to be a multi-faceted concept (Rempel, Holmes, & Zanna, 1985). In parallel to the theoretical contributions from a sociological perspective, research on human operators’ trust in automation has been conducted in the human factors community. Based on the definitions from a sociological perspective, two essential studies investigated the role of human trust in continuous process control environments and produced results generally consistent with those from the sociological study of trust. These studies (Lee & Moray, 1992; Muir & Moray, 1996) examined the role of operators’ trust in automated systems. Both studies found that operators’ trust was related closely to their utilization of automated systems.

Although the characteristics that impact human trust have been documented in sociological as well as human factors research (Muir, 1987; Rempel et al. 1985), there has been little research regarding decision-making performance in relation to operators trust in automated decision-aiding systems. To address this gap, the research conducted here examined the effect of imperfect automated decision-aids on human decision-making performance along with trust in the automated systems. In particular, the main research focuses were to manipulate factors which have been previously hypothesized to impact trust, and to investigate the effect of these factors on operators' judgment performance, as well as their trust.

Research on people's judgment and decision-making behaviors is very rich in both theoretical and empirical studies. Many decision-making models have been developed to describe people's behaviors from theoretically driven and normative decision-making models like the expected utility model to practically driven and descriptive models, such as recognition-primed decision-making model. Among many models which exist today, the Lens Model framework with its extensions (Cooksey, 1996) includes concepts found in classical decision models (how humans sample and combine information), as well as describing the relationship between available cues and the true state of the environment. Among the modeling extensions within the Lens Model framework, two systems: n-system and hierarchical system models are relevant to the environment of interest in this study: human decision-making supported by automated decision-aids. To capture operators' judgment policies, a Hybrid Lens Model was developed to represent operators' judgments, outputs from an automated system, and operators' reliance on the automated output. One important relationship that must be identified in this study to understand the relationship between operators' trust in automated decision-aids and utilization of the system is the effect of automated decision-aid's estimates on operators' judgment policy. The Hybrid Lens Model developed here is used to provide such information.

Additionally, this research studied a design intervention intended to help calibrate operators' trust in automated systems. Two research streams were considered and linked to identify a way to support operators' judgment and further calibration of trust. Previous research on people's judgment and decision-making claimed that providing the task information improved people's judgment performance most. According to Balzer, Sulsky, Hammer, Sumner, Birchenough, Martens, & Raymark (1992), people performed a judgment task when they were provided with the task information including a statistical relationship between the environmental criterion and the informational cues, and a relationship between predicted and actual environmental criterion, rather than when they were provided with simple performance feedback. From the list of potential characteristics of trust suggested by Sheridan (1988), understandability or transparency were chosen to be useful in both increasing human operators' judgment performance and calibration of trust. Therefore linking these above two, the cognitive information of automated decision- aid for the task information was used to accomplish both purposes.

This final report is organized as follows:

First, Chapter 2 reviews literature from three different perspectives on human judgment with, and trust in, automated systems. Description of early studies, i.e., Fitts' (1951) list of function allocation between human and machine or automation, are provided to find connections to later studies on automation taxonomies (Parasuraman, Sheridan, & Wickens, 2000; Endsley & Kaber, 1999). Then, trust in interpersonal relationships from the sociological perspective is described to form the basis of operators' trust in automated systems from a human factors perspective. An emphasis is placed on Sheridan's (1988) description of the definition of trust characteristics, identifying several experimental variables to be controlled. The environment where human operators are supported with automated decision-aids is described to discuss several important relationships that must be identified to understand human-machine interaction. Finally, the human trust in virtual teams is discussed with the emphasis placed on multi-agent systems, including automated decision-aids and human operators.

In Chapter 3, a detailed description of the Lens Model framework with its extensions (Cooksey, 1996) is provided to model a variety of relationships among the components, which are described in the previous chapter. The conceptual relationships between the human operators and the automated decision-aid are translated into the parameters within the Lens Model framework. Not only was the separate Lens Model of each component in this study provided, but also a description of the Hybrid Lens Model developed is given to understand the relationships. A Hybrid Lens Model describes a model of human operators' judgment with automated decision-aids based on the Lens Model framework. The Lens Model framework provides: useful methods to capture operators' judgment policy and performance, the important relationships of capturing the effects of the decision-aids on the operators' judgment, reliance on automated decision-aids, and a brief discussion on statistical concepts.

Chapter 4 presents the independent variables to instantiate the effect of various characteristics of automated decision-aids on human judgment performance and its relationship with operators' trust in such systems. The characteristics discussed in the previous chapter are experimentally controlled. This chapter also presents the dependent measures that must be collected to evaluate and understand operators' judgment behavior with automated decision-aid. Finally, the experimental framework developed to investigate human trust in automated decision-aids is discussed.

Additionally, Appendices A-C show the lists of literature related to human trust in a variety of settings: Appendix A contains a list of literature with human trust in virtual teams, Appendix B contain lists and summaries of literature with human trust in a sociological environment, and Appendix C contains lists and summaries with trust in a variety of environments.

2.0 LITERATURE REVIEW

2.1 Overview

Understanding human operators' judgment and decision-making behavior with automated decision-aids is important in order to support and increase their judgment performance. There are many possible ways to support human decision-making activities, such as designing information displays which are easy to understand, devising training programs, or implementing automated decision-aids.

The goals of this chapter are four fold. First is to look at how human operators' judgment and decision-making performance can be affected by imperfect automation technology. Thus, it is necessary to review literature on human centered automation. Specifically, the extent to which human operators are considered in the design and study of automation technology needs to be reviewed.

Second, there are clear differences between automated systems in supervisory control contexts and automated decision-aids, which are generally designed to support decision-making activities rather than actually controlling the systems. Thus, a review of research on automated systems and automated decision-aids will be discussed, specifically from the human centered automation perspective. These areas of research will pave the ground for the discussion on human trust in automation, especially human trust in virtual environments.

Third is the issue of operators' trust in automation. Research on operators' trust in automation is a relatively new area, in which research has been performed primarily in the context of automated systems in supervisory control. A review of operators' trust in automation will be provided, leading to an analysis of judgment performance and trust in automated decision-aids.

Fourth, human decision-making has been an important issue of study for many years and researchers have approached the problem of understanding human decision-making from a variety of perspectives. Although there are many decision-making modeling frameworks to date, these models have not been used to represent the situations where human operators are supported with automated decision-aids, which provide estimates of the environmental states. Therefore, an effort to develop a modeling design based on Brunswik's Lens Model framework (Brunswik, 1952; Hammond, Stewart, Brehmer, & Steinmann, 1975) will be described.

2.2 Background Literature

2.2.1 Human Centered Automation

2.2.1.1 Concepts

Automation technology and human operators are both components of complex systems. Many different problems have been documented when humans are not considered at the design stage of automation, such as new forms error (Sarter & Woods,

1995), increased mental load (Wiener, 1988), increased monitoring demand (Parasuraman, Molloy, & Singh, 1993), trust problems (Lee & Moray, 1992) decision biases (Mosier & Skitka, 1996), skill degradation (Hopkins, 1994), out-of-the-loop performance (Wickens, 1992), low level of situation awareness (Endsley & Kiris, 1995; Wiener & Curry, 1980), cognitive overload (Kirlik, 1993), and masking of incompetence (Hopkins, 1994). Four characteristics can be summarized as important aspects to consider in designing automated systems: complexity (Sheridan, 1988), autonomy (Sarter & Woods, 1995), the coupling or interdependence (Woods, 1988), and opacity (Woods, 1988).

First, complexity refers to the interaction between and the transitions across a large number of different system elements or modes. It is difficult for human operators to understand systems that are very complex in nature. Such understanding is essential to facilitate building or maintaining operators' mental models of the systems' behavior.

Second, autonomy or authority refers to automation taking over control of a variety of functions, such as decision-making, without notifying its decisions to human operators. "Silent" automation indicates the system that provides none or poor feedback about its activities or intentions (Norman, 1990). While autonomous automated systems refer to systems which can function for long periods of time without further intervention (Woods, 1993), a degree of authority indicates that the systems are also capable of taking over control of the monitored process from another agent if it decides that intervention is warranted based on its perception of the situation and its internal criteria (Sarter & Woods, 1995). Increasing autonomy and authority creates new monitoring and definitions of roles of the components within a system. Human operators as supervisors have to keep track of the status and changes of modes (Sarter & Woods, 1995).

Additionally, the issue of autonomy brings an issue of responsibility (Mockler, 1989; Moon & Nass, 1998). As Zeide & Liebowitz (1987) noted, the current judicial system indicates that the liability must be traced back to human resource. Diffusion of responsibility is more likely when malfunctions happen. Especially for those systems where human operators are responsible for system effectiveness and safety, human centered automation must remain firmly in command of all components of the system.

Third, the coupling or interdependence refers to the degree of dependence between parts within a system. Coupling is a related property that refers to a system design where a particular input results not only in a particular expected and desired behavior but also has additional 'indirect' side effects (Woods, 1988). The stronger the interdependence or coupling, the more parts are intertwined in terms of individual functions. This creates the problem of latent failures (Reason, 1990), meaning the negative effect of one part of the system has consequences that are not revealed or activated until some other enabling conditions are met. Not only may this make providing necessary information for diagnosing situations difficult (i.e., fault diagnosis), but also increases the demand on human operators' mental model of automation functioning or connectivity. Without such understanding, humans cannot form clear mental models of how the automated systems should behave; therefore they cannot predict the automated

systems behavior in the variety of circumstances they may encounter. That is, highly coupled automated systems demand better trained operators who must have a better understanding of the automated systems' capability or functioning (Woods, 1996). However, this defeats the original purpose of implementing automation from the beginning.

The above three characteristics lead to the final characteristic: the degree of transparency (Wickens, 1992). High levels of complexity, of very silent automation, and of interdependence eventually make the automated systems opaque, preventing the operators from observing the inner workings of the systems. Although the display format is a way to resolve this issue of transparency, it would be much better to refer to interface and feedback issues. This term captures the fundamental relationship among the components of a system, the environment and human operators, and context of observation.

2.2.1.2 Frameworks for Human Automation Systems

One method that has been applied to the design of human centered automation is the use of frameworks as taxonomies that address different styles of function allocation between humans and automated systems. For example, Fitts (1951), shown in Table 1, provided a list of disciplines when functions must be allocated between human and machine. Basically, this list is based on the superiority concept in that those tasks that humans do better must be assigned to human operators and vice versa. For example, machines are better at entertaining the same judgment tasks repeatedly based on fixed or certain criteria while human operators are better at making judgments based on inductive reasoning.

Table 1. The Fitts List (Fitts, 1951)

Men (humans) are better at:
• Detecting small amounts of visual, auditory, or chemical energy
• Perceiving patterns of light or sound
• Improving and using flexible procedures
• Storing information for long periods of time and recalling appropriate parts
• Reasoning inductively
• Exercising judgment
Machines are better at:
• Responding quickly to control signals
• Applying great force smoothly and precisely
• Storing information briefly, erasing it completely

- Reasoning deductively
- Doing many complex operations at once

This list provided a starting point to approach allocating functions between human operators and machines, which further allow the designers of automated systems to make decisions on the types of automated systems. However, later frameworks addressed more explicit aspects of shared and dynamic function allocation with more complex automated systems.

Sheridan and his colleagues (Sheridan, 1988; Sheridan & Verplank, 1978), proposed a somewhat different kind of list: a scale of degrees to which automated systems are involved with human operators' activities, such as generating decision alternatives, decision-making, and action implementation, shown in Table 2. According to Sheridan's (1992) definition, automation is "the automatically controlled operation of an apparatus, a process, or a system by mechanical or electronic devices that take the place of human organs of observation, decision, and effort." This definition of automation can include a basic form of automation such as a simple mechanical device, to a more complex form of automation that can generate advanced and processed data for rendering human operator's judgment tasks easier.

Table 2. Sheridan's (1988) Levels of Automation of Decision and Action Selection; Levels Associated with Automated Decision-aids are Annotated

	Level	Description	
High	10	Complete autonomous computer decision and action	
	9	The computer informs the human only if it decides to	
	8	If asked, the computer provides information to the human	
	7	Automatic execution, then necessarily informs the human	
	6	Automatic execution after a certain amount of time	
	5	Execution with the human approval	Automated
	4	The computer suggests one alternative	Decision
	3	The computer narrows down to several alternatives	Aids
Low	2	Human given alternatives of decision and action	
	1	Complete human responsibility in decision and action	

The levels of automation suggested by Sheridan (1988) show that the number of functions of automated systems gradually increases from a low to a high level of

automation and so the role of such systems becomes more critical in performing the tasks. That is, this list shows less involvement of human operators as the level of automation increases. The highest level of automation indicates a total supervisory control where the operators may not have the authority to override the automated systems' outcomes or actions. On the other hand, the lowest level of automation indicates the complete responsibility of human operators for all tasks.

Factors such as the human operators' cognitive capability, and the reliability of automated systems, need to be considered in selecting an appropriate level of automation for a situation. For example, there can be cases where adequate performance on critical or important tasks cannot be met with automation technology. In this case, the level of automation must be reduced, and accuracy weighted more heavily than advantages of automation such as speed.

An approach taken by Endsley & Kaber (1999), shown in Table 3, developed taxonomy for levels of automation. The levels of automation are similar to Sheridan's list of automation. In any system, in which human operators are assisted with automated systems, there exists four generic functions that must be assigned to either human operators or machines: (1) monitoring – scanning displays to perceive and identify system status; (2) generating – establishing alternatives or options to achieve purposes; (3) selecting – making decisions on a particular strategy; and (4) implementing – implementing the decision into necessary actions. Similar to Sheridan (1988), ten levels of automation were identified by varying degrees of involvement of automation in performing tasks (Tbl. 3) are discussed below.

Table 3. Level of Automation Adapted from Endsley & Kaber (1999)

Level of automation		**Roles**			
		M	G	S	I
1	Manual control	H	H	H	H
2	Action support	H/C	H	H	H/C
3	Batch processing	H/C	H	H	C
4	Shared control	H/C	H/C	H	H/C
5	Decision support	H/C	H/C	H	C
6	Blended decision-making	H/C	H/C	H/C	C
7	Rigid system	H/C	C	H	C
8	Automated decision-making	H/C	H/C	C	C
9	Supervisory control	H/C	C	C	C
10	Full automation	C	C	C	C

M: Monitoring, G: Generating, S: Selecting, I: Implementing, H: Human, C: Computer

(1) Manual control: At this level, a human performs all four tasks including monitoring the status of the system, generating alternative decisions, making decisions, and finally implementing the decisions.
(2) Action support: Computer or machine shares or takes over monitoring and implementing activities, meaning that the system assists the human operator with the performance of the chosen actions.
(3) Batch processing: The automated system carries out the human's decision-making, which is selected among alternatives that are also generated by the human. Simple automated systems exist such as batch processing in manufacturing operations.
(4) Shared control: Among the four functions defined previously, selecting an alternative remains with the human operators. Shared control means that both the system and the human share control over the action implementation.
(5) Decision support: Decision support indicates that the system support generates decision alternatives for the human to select from. However, it must be noted that the human can also generate other or same alternatives, which further allow the human to compare it against those from the automated system. As Endsley & Kaber (1999) noted, this is representative of many expert systems that guide the human through the problem solving process. The difference from the previously described level of automation, shared control lies on the focus of action implementation. At this level, actions are automatically implemented without any human involvement.
(6) Blended decision-making: This is equivalent to level five in the Sheridan (1988) taxonomy of automation in which the automated system executes the human's decision only with approval. As is in the previous level, the system executes the decision automatically.
(7) Rigid system: Rigid system is representative of a system that only a limited number of alternatives are generated by the system. This indicates that the human may or may not have a chance to generate other options.
(8) Automated decision-making: Generation of alternatives and making decisions are entertained by the automated systems as well as implementing the selected option. An intermediate stage is possible where the list of alternatives can be augmented by the human operator. However, the selection process still remains at the automated system.
(9) Supervisory control: This is the level of automation emphasizing the human operators' activity that monitors and supervises what the automated system is performing. In a highly automated system, the human operator monitors and intervenes with the process where necessary.
(10) Full automation: At this level, the automated system performs every task necessary. Therefore, the human operator is out of the loop and not is allowed to intervene with the process.

Crossing the level of automation with the four functions within human-machine interaction, Endsley & Kaber (1999) defined an assignment scheme between humans and automated systems for each level of automation. Allocating these four functions to either human operators or machines, depending on the level of automation, they conducted an experiment to investigate whether different levels or types of automation might have an effect on operators' performance. They found that levels of automation which distributed the roles of option generation and implementation between human and/or computer

servers had a significant impact on overall system performance. Specifically, the level of automation that combined human generation of options with computer implementation produced superior overall performance during normal operations, as compared to purely manual control and to higher levels of automation involving the computer generation of options.

Reviews of two important studies on the levels of automation are provided. It is apparent that the two taxonomies presented by Sheridan (1988) and Endsley & Kaber (1999) are similar in terms of definitions of the functions of the humans and the automated systems. The comparison is shown in Table 4.

Table 4. Comparison of the Levels of Automation between Sheridan's (1988) and Endsley & Kaber (1999)

Sheridan (1988)	LOA	Endsley and Kaber (1999)
Complete autonomous computers	High ↑	Full automation
The computers informs the human only if it decides to		Supervisory control
If asked, the computer provides information to the human		Supervisory control
Automatic execution, then necessarily informs the human		Supervisory control, Automated decision-making
Automatic execution after a certain amount of time		
Execution with the human approval		Rigid system, Blended DM
The computer suggests one alternative		Blended decision-making
The computer narrows down to several alternatives		Decision support
Human given alternatives of decision and action		Batch processing, Action support
Complete human responsibility in decision and action	↓ Low	Manual control

Although both taxonomies are very similar and focus on responsibility of the decision-making task and implementing the decisions, Sheridan’s (1988) level of automation is more focused on how information flows between the two parts of the system. For example, four levels are related to the presence of decision-making notification and the timing of notification. On the other hand, Endsley & Kaber (1999) focused on the assignment of the four functions that both agents might take part in controlling.

Parasuraman, Sheridan, & Wickens (2000) devised a simple model for types and levels of human interaction with automation. When designing an automated system, the designer or the decision-maker has to make many decisions such as what to automate and how far the system must be automated based on various aspects, such as the operators' mental resources, or a cost and benefit analysis based on the degree of criticality or consequences from which mis-operations or malfunctions may cause. According to Parasuraman, Sheridan, & Wickens (2000), there are several factors to consider when deciding what or how to automate systems, by defining types or levels of automation, respectively.

According to Parasuraman, Sheridan, & Wickens (2000), there are four types of tasks relevant for automated systems: Information Acquisition, Information Analysis, Decision Selection, and Action Implementation. They define information acquisition as the sensing and registration of input data. Information analysis involves cognitive functions such as working memory and inferential processes. Examples are the predictor displays (Morphew & Wickens, 1998) for the cockpit and trend displays in process control (Moray, 1997). Decision selection involves varying levels of augmentation or replacement of human selection of decision options with machine decision-making, while action implementation refers to the actual execution of the action choice.

Among many factors affecting these two questions of types of and levels of automation according to Parasuraman, Sheridan, & Wickens (2000), automation reliability and cost of decision outcomes, in terms of risk involved, were identified as the most critical factors. For both information acquisition and analysis types of tasks, they suggest a high level of automation if the associated systems are reliable, i.e., detecting signals (system B in Fig. 1). This is especially the case of where human operators do not have access to the remote sources of information. Therefore, the main factor which impacts the level of automation in these types of tasks is the reliability of the available automated systems. Additionally, feasibility of automation becomes a decisive factor in determining whether these devices should be automated or not. However, there are many cases where it is simply impossible for human operators to acquire and analyze the information, such as in the air traffic control system (ATC) that Parasuraman, Sheridan, & Wickens (2000) used. For example, in the ATC system, it is not feasible for human operators to measure the necessary information, such as altitude of an airplane. Additionally, the level of automation for the decision selection type of automation should be low if it involves a high level of risk. In a military target identification environment where human operators must identify unknown objects based on measured or sensed information, the risk function involved tends to be very high. Therefore, the final decision selection must remain at the human operators' discretion. Figure 1, shows the Parasuraman, Sheridan, & Wickens' (2000) framework of the levels and types of automation.

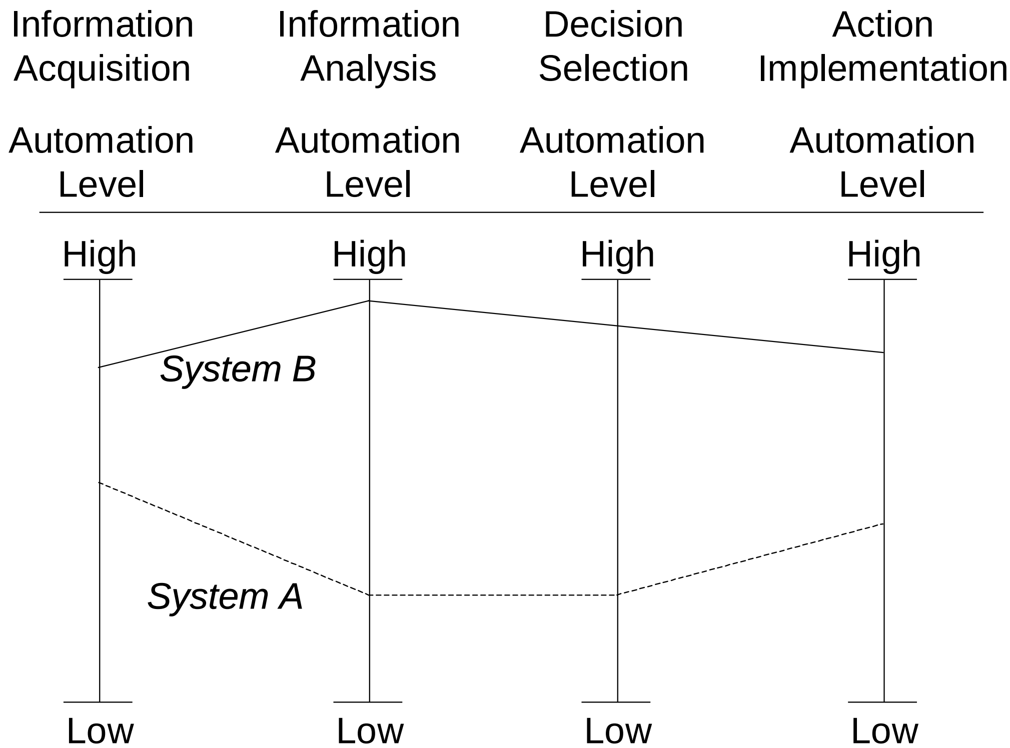

Figure 1: Types of and Levels of Automation (Parasuraman, Sheridan, & Wickens, 2000)

Based on the literature review, it seems that there is common ground among the three studies conducted about the types of or levels of automation. Combining Parasuraman, Sheridan, & Wickens' (2000) theoretical study on the levels of automation affected by a variety of factors, with Endsley & Kaber's (1999) empirical results that showed the best performance when operators generate options and machines implement them, results in a recommended set of levels represented on Parasuraman, Sheridan, & Wickens' (2000) taxonomy of levels of automation, shown in Figure 2.

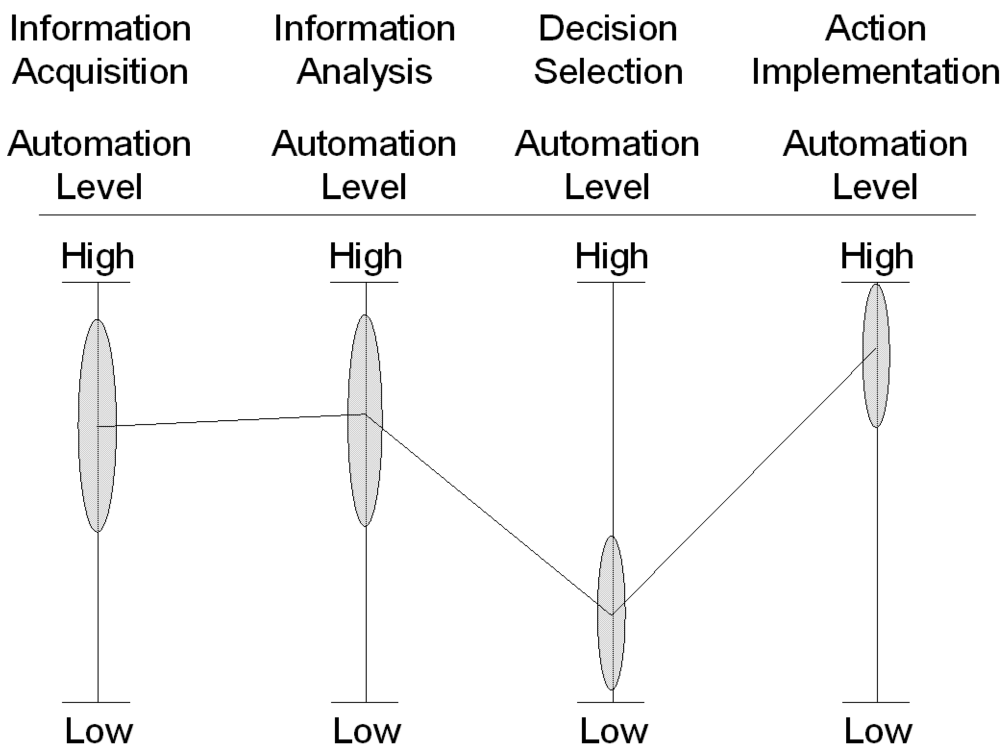

Figure 2: Levels of Automation that Produced Superior Overall Performance in Endsley & Kaber (1999) Superimposed on Parasuraman, Sheridan, & Wickens' (2000) Levels of Automation

2.2.1.3 Summary

Beginning with the Fitts (1951) list, the role of machines or computers has been scrutinized to assist human operators to better perform interactive tasks with the environment via automated systems. Advancing from the Fitts' philosophy of allocation based on superiority, Sheridan (1988) proposed a level of automation that gradually varies the degree of the human operators' involvement in performing tasks, starting from manual control in which every task must be performed by human operators, to completely automation-controlled systems in which human operators monitor the automated systems. The main focus of this list lies on the locus of authority for decision-making and the execution of the decisions. A similar approach taken by Ensley & Kaber (1999) developed a taxonomy to describe the role of two agents in performing tasks. Their results showed that levels of automation distributing the roles of option generation and implementation between human and/or computer servers had a significant impact on automated system performance.

Additionally, Parasuraman, Sheridan, & Wickens (2000) proposed another approach varying in two dimensions; levels of and types of automation. Simply varying from high to low level, the levels of automation are defined as a continuum. They

identified four types of automation, of which could vary in the degrees of automation; information acquisition, information analysis, decision selection, and action implementation. The types of automation can be considered sections of physical components within an entire system. Parasuraman, Sheridan, & Wickens (2000) also described the critical factors such as reliability, or cost of decision outcome, affecting the level of and the type of automation.

In summary, several studies have been discussed to describe approaches to designing automated systems. Whether it is levels of automation or types of automation, it is apparent that the attention of importance of human operators in designing automated systems has been substantially increased.

2.2.2 Research on Automated Decision-aids

One type of automated system that generates alternatives about the state-of-the-environment or possible courses of action can be called automated decision-aids. Automated decision-aids refer to the automated systems that support human operators in problem solving activities or selection of decisions, which require an embedded knowledge base. According to Parasuraman, Sheridan, & Wickens' (2000), automated decision-aids refer to automation in decision generation and selection. In this section, research on automated decision-aids will be reviewed. The definitions of automated decision-aids, research findings on such systems, and relations of this literature to the automated decision-aid in the present work will be discussed.

2.2.2.1 Introduction and Definition of Automated Decision-aids

As decision-making environments become more complex and data intensive, the use of automated decision-aids is likely to become even more commonplace and more critical to the process. The benefits of automated decision-aids and expert systems in terms of increased efficiency for data monitoring and analysis capabilities are fairly obvious. However, the behavioral influences of decision-aid use must be assessed to determine the benefits, limitations, and costs associated with decision-aid development and implementation.

Research has shown that people are very limited in their ability to process information in uncertain environments, especially so with judgment and decision-making (e.g., Tversky & Kahneman, 1974). Researchers from various disciplines have demonstrated that decision-aids can assist human judges to make more accurate and consistent judgments (Morrison, Kelly, Moore, & Hutchins, 1998). Enhanced judgment performance can be attributed to incremental information processing capacity, or compensating for shortcomings of human judgments, such as biases and anchors.

There are many different kinds of automated decision-aids that have been designed to support humans in understanding the environment, diagnosing the situation, or engaging in problem solving activities (Rasmussen, Pejtersen, & Goodstein, 1994). For instance, in a system where human operators do not have any access to the actual state of the environment, information regarding the state of the environment is usually

collected via various methods such as sensors or other personnel. It is the automated decision-aid's job to integrate the information and provide an estimate of the state of the environment. That is, automated decision-aids are defined as providing an estimate of the state of the environment based on the information captured from the environment in order to support human operators' judgment and decision-making (Morrison et al., 1998). A typical example would be a human operator identifying an unknown object as friendly or hostile, based on the information collected via sensors, along with an identification estimate provided by the automated decision-aid. Together with the informational cues, which are measured or sensed from the state of the environment, the estimate can provide valuable information about the environmental state for the human operator making the final judgment.

2.2.2.2 Research Findings on Automated Decision-aids

Although the importance of automated decision-aids in assisting human operators' judgment and decision-making tasks has been emphasized over decades, little work has been done empirically in terms of how automated decision-aids must be designed to incorporate human operators. The leading force in designing and implementing automated decision-aids has been current technology rather than focusing on how people interact with and react to the automated decision-aids.

Will (1991) investigated false dependence on technology and performed an experiment with reservoir petroleum engineers in which misleading suggestions were provided about the appropriate methodology to address the given problem situation. The expert information system used by many engineers was to support them in the analysis of pressure buildup problems in backward-chaining by providing both "how" and "why" explanations. Falsified statements in fundamental interpretation of principles in problem solving were provided to investigate their impact on the decision confidence and performance depending on the level of expertise. It is unknown however to what degree the statements from the expert system were falsified from the true solutions. He found no significant difference in decision confidence and concluded that experts, in general, did not identify errors in the expert system. This result is somewhat consistent with Riley's (1996) in that no significant difference was found between the pilots and college student in activating automation. He also found that novices consistently attributed significantly more to system success than experts, meaning that experts did not view the expert system as helpful in contributing to better job performance or decision-making satisfaction.

One factor that can affect operators' decision-making performance is the complexity of environment. It is obvious that the more complex the environment is, the more difficult the judgment tasks become. Similarly, the complexity of automated decision-aids can have an impact on operators' understanding of such systems. For example, automated decision-aids for the highly structured environment need not to be complex expert systems that require a variety of knowledge built-in (Rasmussen et al., 1994). Simple deterministic types of automated decision-aids can perform the task efficiently. Glover, Prawitt, & Spilker (1997) investigated whether the decision-aid could have an effect on users' acquisition of knowledge and their reliance on a decision-aid. They performed an experiment with a computerized determination of tax liabilities,

including computation of taxes on net capital gains. The decision-aid used in this study, similar to the actual worksheets used by the government, represents a structured task in which participants could be supported by answering initial questions regarding the status of the tax reporter. The decision-aid is more or less a decision template which users have to collect necessary information and enter into the template. Then, the decision-aid provided a suggested solution for the users. Therefore, the information collection and analysis tasks remained with the users to figure out what information to look for and how to draw conclusions from the collected information. Their results were consistent with prior research, indicating that the use of decision-aids enhanced performance.

On the other hand, highly unstructured problem domains, where events are intertwined and multiple causes can create multiple stages of events, need not only expert systems with specific knowledge databases but also human experts. Morrison et al. (1998) developed a decision-aid to support human operators in understanding environmental situations, making judgments and planning or employing countermeasures in a complex environment. The information provided in this decision support system was required to make judgments about the environmental state and to evaluate the degree of risk or threat so they could prioritize different activities. For example, the decision support system provided an assessment as to whether or not an unknown object the human operator intended to identify was a threat and provided supporting evidence, counter evidence, and assumptions the assessment was based on. Morrison et al. (1998) performed an experiment evaluating the effectiveness of this decision support system and found that the operators reported the system useful for making judgments, and the operators were able to recognize the critical tracks earlier than without the decision support system. Additionally, they found that the operators were able to take more timely and appropriate defensive actions against imminent threats.

In addition to some benefits in terms of performance, it has also been argued that decision-aids can facilitate task-related learning (e.g., Libby & Luft, 1993). This argument is based on the claim that structural relationships embedded in the environment can be transferred via the decision-aids to the human judges, which can provide useful information about the environment to establish appropriate knowledge.

When automation is involved in the decision-making process, human decision-makers may be negatively impacted in terms of the acquisition of knowledge. Glover et al. (1997) hypothesized that users with decision-aids would acquire less knowledge compared to those without decision-aids. The decision-aid used by Glover et al. (1997) was a simple and highly structured decision-aid that provided a suggested solution for a problem in a simple calculation of tax liability. They found that the knowledge score of participants, measured by asking them to write about the domain concept, was significantly less for participants with the decision-aid than those without the decision-aid. This result was in line with arguments by Wiener (1980), who claimed that one of the negative impacts of automation was operators' skill degradation.

Another important issue to consider regarding automated decision-aids is the method to present the outcomes. There can be many ways to simply present information,

i.e., graphically or numerically. In this regard, many studies showed conflicting results on the types of information displays on a variety of application areas. Grounds & Wiley (2001) conducted an experiment to investigate the effect of different display interface designs on operators' trust in the decision-aid. They found no significant effect of different display types (configural display, bar graph, and alphanumeric) on operators' decision time to agree or disagree with the recommendations. On the other hand, several other researchers have found such an effect (Crocoll & Coury, 1990; Lee, Gore, & Campbell, 1999; Yeh & Wickens, 2001). For example, Yeh and Wickens (2001) performed an experiment focusing on the effect of providing intelligent cuing information to guide attention to targets and reliability of the automated attention-directing device. They found that the presentation of cuing significantly improve the detection accuracy of low-salience objects. However, the presentation did not help participants in detecting accuracy of expected, high-salience targets. Also, they found that the unexpected target was less likely to be detected when it was presented with a cued object rather than with an uncued one.

2.2.2.3 Automated Decision-aids in the Current Work

In this research, an automated decision-aid typically refers to the types of automation ranging from information acquisition to information analysis according to Parasuraman, Sheridan, & Wickens (2000), which they collectively called information automation. In such systems, data or information collected from the environment is analyzed and fused together to produce further advanced or enhanced information to be useful for human operators making final judgments. It could also be categorized as the decision selection type of automation because of the estimates generated by the automated decision-aid. For example, the automated decision-aid notifies the human operators of the decision outcomes when it decides to, or upon human operators' request. Note that the human operators are not involved in the decision- making process. Because of this reason, the automated decision-aids in this study are limited to the information automation in which the decision selection is allocated to the human operators.

Specifically, automated decision-aids in the current work refers to systems supporting human operators' decision-making activities by integrating sensed information from the environment, which the operators do not have any access to, and providing sensible estimates of the state of the environment based on a computational algorithm. Therefore, the automated decision-aid collects necessary data or information representing the environment and combines them into an estimate in a more comprehensible form. Therefore, automated decision-aids support human operators' computation and information integration task which is otherwise cognitively challenging task. Each set of information is probabilistically related to the environment which means that the information is not completely diagnostic of the environment, and that the information fusion mechanism plays a critical role in generating valid environmental estimates. This is important because it not only affects the quality of the estimates, but also it has an impact on operators' trust in automated decision-aids. Consequently, it may further impact utilization of the automated decision-aid.

As Endsley & Kaber (1999) described, level 5 automation decision support, indicates that the computer generates a list of decision options that the human can select from or the operator may generate his or her own options. This level represents of many expert systems or decision support systems that provide options guidance which the human operator may use or ignore in performing a task. The automated decision-aids in this current work can be categorized as the level 5 in Endsley & Kaber's (1999) taxonomy of the level of automation, and can be any levels between level 2 and level 5 in Sheridan's level of automation. In fact, the level of automation that Endsley & Kaber (1999) described as producing superior overall performance is the same level as the level of automation in this study.

2.2.2.4 Summary

Automated decision-aids are defined here as systems which provide an estimate of the state of the environment based on the information captured from the environment, in order to support human operators' judgment and decision-making. Rather than controlling systems, automated decision-aids support operators in terms of collecting information, integrating the sets of information into more comprehensible form of alternatives about the environmental states to ultimately support the decision-making task. The main support is to support human operators' cognitive information integration task.

Previous research on the levels of and types of automation indicated that automated decision-aids are high in terms of the level of automation at the information acquisition and analysis, and are low at the decision selection. The automated decision-aids in the current work were mapped into two studies on the levels and types of automation. The automated decision-aid in this current study follows the general definition of automated decision-aids, which human operators are still responsible for the final decision-making task based on the sets of environmental data and the environmental estimate from the automated decision-aid.

Finally, several important factors were described that might have influence on design decisions on the level and types of automation. Careful consideration of these factors must be exerted to provide firm support to human operators interacting with uncertain environments.

2.2.3 Research on Trust in Automated Systems

Needless to say, trust in interpersonal relationships is a key factor in building a good relationship. Such concepts have suggested to many researchers that trust in automated systems may play a key role in the "relationship" between humans and automation. Based on the sociological perspective of people's trust in interpersonal relationships, several studies have been conducted to investigate the role of operators trust in the use of automated systems. It is necessary to review the studies on people's trust from the sociological perspective, which will pave the way to look at several studies from the human factors perspective.

2.2.3.1 Research and Models of Trust from the Sociological Viewpoint

Trust is a key component in developing interpersonal relationships. A certain level of trust in others can make interpersonal relationships successful. Trust is more than simple agreement or disagreement between two parties, although accumulation of these interactions can certainly evolve into developing trust or distrust.

There have been several studies investigating the importance of trust on interpersonal relationships, different aspects of people's trust in interpersonal environments and how they develop their trust in others.

Barber (1983) proposed a definition, which includes multidimensional aspects and necessary attributes of trust in terms of three specific expectations: persistence, technical competence of role performance, and fiduciary obligations and responsibilities. According to Barber, persistence of natural and moral laws provides a foundation of trust by establishing constancy in the fundamental moral and natural laws. These expectations provide the basic conditions for social and physical interactions. Technically competent performance, on the other hand, supports expectations of future performance based on capabilities, knowledge, and expertise. Barber discussed this dimension of trust referring to the ability of the other partner to produce consistent and desirable performance which can be subdivided to include three types of expertise; everyday routine performance, technical facility, and expert knowledge. These types will be explained later in conjunction with another human factors error nomenclature.

Barber's third dimension of trust, fiduciary responsibility, concerns the expectation that people have moral and social obligations to hold the interests of others above their own. Fiduciary responsibility extends the idea of trust beyond that based on performance to one based on moral obligations and intentions. This dimension becomes important when agents cannot be evaluated because their expertise is not understood, or in unforeseen situations where performance cannot be predicted. Here expectations depend upon an assessment of the intentions and motivations of the partner, rather than past performance or perceived capabilities.

In addition, the dimensions of trust proposed by Barber (1983) and Rempel et al. (1985) emphasized not only components of interpersonal trust, but also the dynamic characteristics of trust toward a partner, regarding trust as a generalized expectation related to the subjective probability an individual assigns to the occurrence of some set of future events. Rempel et al. (1985) found people's trust has various aspects that can be developed over time. They defined the term as a multifaceted important intermediate factor in a successful interpersonal relationship. The characteristics of trust they investigated are predictability, dependability, and faith. These are defined both as the characteristics of trust in an interpersonal environment and as the developmental stages over time, faith being the final stage, which is usually shown in a very mature relationship. According to Rempel et al. (1985), predictability, which represents the consistency of recurrent behavior and the stability of the social environment, forms the basis of trust early in a relationship. As interpersonal relationships progresses with further

experience, dependability, which represents a more common understanding of the stable dispositions, becomes an important basis of trust between humans and focuses on an evaluation of the qualities and characteristics attributed to the partner. In this stage therefore, the centroid swings and shifts away from the evaluation of the partner's dispositional attributes to the partner. Faith, on the other hand, describes the aspects of trust or belief that must go beyond the available evidence to accept a given supposition as truth. Their survey results showed that these characteristics are developed in the order as the relationships mature.

Results from other studies (e.g., Barber, 1983; Rempel et al. 1985) are generally in consensus in that people develop their trust gradually based on the evidences, and once betrayed, take long time to recover.

2.2.3.2 Frameworks of Trust in Human-Machine Systems

Based on the sociological definitions of trust described in the previous section, Muir constructed a model of human trust in automation by incorporating the dimensions of trust proposed by Barber and three more dimensions from Rempel et al. (1985). According to Muir's interpretation, all three of Barber's (1983) meanings of trust seem applicable to the human-machine relationship and become a basis for the framework. Thus, Muir's work forms the first "transition" of the concepts of trust in human-human relationships to human-automation relationships. Muir's work in fact suggests that this is an identity-transform, i.e., that the human-human trust concept extends directly to the human-automation relationship case.

Muir (1994) also identified one of Barber's aspects of trust, technical competent performance, with Rasmussen's (1983) taxonomy of behavior; skill-, rule-, and knowledge-based behavior. Interpreting Rempel et al. (1985) model as a hierarchical stage model as being able to account for changes in operators' trust as a result of experience on a system, where trust develops over time, Muir produced a framework by crossing Barber's (1983) dimensions of trust, with Rempel et al. (1985) framework; this crossed relationship among these various dimensions is shown in Table 5. Table 5 shows an integrated model of trust created by Muir (1994) by crossing Barber's (1983) multidimensional characteristics of trust and Rempel et al.'s (1985) model of the dynamics of trust to investigate the meaning and the role of trust in human-machine relationships.

Table 5. Muir's Framework of Trust by Crossing Barber's (1983) Taxonomy of Trust and Rempel et al. (1985) Model of the Dynamics of Trust, from Muir (1994)

	Basis of expectation at different levels of experience		
Expectation	Predictability	Dependability	Faith
Persistence			
Natural physical	Events conform to natural laws	Nature is lawful	Natural laws are constant
Natural biological	Human life has survived	Human survival is lawful	Human life will survive
Moral social	Humans and computers act 'decently'	Humans and computers are 'good' and 'decent' by nature	Humans and computers will continue to be 'good' and 'decent' in the future
Technical competence	*j*'s behavior is predictable	*j* has a dependable nature	*j* will continue to be dependable in the future
Fiduciary responsibility	*j*'s behavior is consistently responsible	*j* has a responsible nature	*j* will continue to be responsible in the future

Based on the definitions of trust in an interpersonal environment, Muir & Moray (1987) applied the same definition to a human machine environment where operators had to control multiple system components which can be switched between manual and automatic operation. Each component could be controlled in manual or in automatic in which the human operator's task in automatic mode of control is simply to monitor the automation.

Additionally, Sheridan (1988) provided a list of attributes of trust specifically for the human machine environment.

<u>Reliability</u>: This factor refers to a system of reliable, predictable, and consistent functioning. Most of the prior definitions of trust addressed this attribute as the first step in developing trust, based on the premise that a person who behaves in a consistent manner will be trusted easily. Lee & Moray's (1994) experiments showed the importance of reliability affecting the level of trust in automation, by manipulating the error magnitude. However, Lee and Moray's definition of reliability is broader than Sheridan's. While the latter indicates an automated system's consistent functioning, the former includes the degree to which automatic controlled values are close to the target values as well.

Validity: The validity of an automated system refers to the degree to which it produces correct output. This implies how faithfully the automated system can estimate the environmental state. In operations in which information is integrated from multiple inputs or sensors to produce estimates of the environmental state, validity refers to the distance between the environmental state and the output from the automated system. It seems intuitive that the automated system that produces more valid outputs to the human operators will be trusted more.

Robustness: Robustness supports expectations of future performance based on capabilities and knowledge not strictly associated with specific circumstances that has occurred before. Sheridan (1988) stated this as "meaning demonstrated or promised ability to perform under a variety circumstances." For example, consider the sudden increase of water inflow into a system; this may cause an automatic valve system to malfunction. If the automated system was designed to handle this situation, whether it is expected or unexpected, it will still provide a useful way to control the situation. This should increase operator's trust in the system.

Glover et al. (1997) performed an experiment investigating the effect of an inferior decision-aid on users' performance in the tax calculation domain. As discussed in the previous section, the decision-aid used in the study was simple, representing a highly structured environment. However, the decision-aid was designed to mainly focus on the "typical" cases. That is, the decision-aid was not robust to consider a variety of cases that can exist in the real world. Their experimental result showed that participants' performance was significantly better with the decision-aid than those without the decision- aid, but only for those typical cases. For non-typical cases, participants without the decision-aid significantly outperformed those with the decision-aid. From the result, it appears that if the decision-aid is not robust enough to handle many different cases, it is of no use in supporting human operators' judgment and decision-making activities.

Transparency: Transparency refers to the extent to which the inner workings or logic of the automated systems are known to human operators, to assist their understanding about the system. Opaque systems become another source of uncertainty that can have negative impact on operators' judgment and decision-making performance. Unless human operators understand how the automated system performs, they may either disregard or totally rely on the automated system without questioning its capability. Therefore, providing a method for the operators to understand the automated systems better can increase the chance they learn to appropriately rely on the automated systems.

Understandability: The operator's understanding of the automation is important for the expectations that make him trust and use the system appropriately. The operator's intervention will be better and timelier when his trust is well calibrated to the actual trustworthiness of the system. The construct of understandability is related to the development of an appropriate mental model. In designing a machine to aid a human operator, understandability is affected by the degree of transparency of the system in which the operator can "see" the underlying mechanisms through the interface. Hanes & Gebhard (1966) conducted a series of studies specifically addressing the question of man-

machine interaction. The task setting was a highly realistic simulation of a Navy Combat Information Center, and their subjects were experienced naval commanders. Data were obtained from officers in a variety of scenarios with common features of a computer aid which had varying degrees of impact on the solution of the problem of defending the simulated ship from simulated air attacks. They found that officers were much more willing to accept the computer recommendation if they agreed with the logic used and/or chose the logic themselves, if they understood the computer program or if the recommendations were unambiguously displayed. From this study along with others, we can see that operators' knowledge about the automated system plays a critical role in determining the reliance on decision-aiding technology.

Another way to increase the level of operators understanding is to develop an appropriate display interface design. The interface design presumably matches the operators' concept of automated systems; the so called mental model. Many studies (e.g., Vicente & Rasmussen, 1992) have claimed that an interface embedded with relationships among the components can enhance operators' understanding, which consequently contributes to an increase in performance. On the other hand, Grounds & Wiley (2001) conducted an experiment to investigate the effect of different display interface design on operators' trust in the decision-aid. They found no significant effect of different display types (configural display, bar graph, and alphanumeric) on operators' decision-time to agree or disagree with the recommendations.

Hanes and Gebhard's (1966) results showed that officers were more willing to accept the computer recommendation if they believed the input data were of good quality, or if the recommendations were unambiguously displayed. It may be that the usefulness of any system depends on the manner in which system information is presented. In the case of a decision-aid which produces the estimate of the environment, there must be a suitable presentation of the uncertainty regarding the system's inferences. This method of presentation becomes the window of opportunity for the operators to reduce the additional uncertainty created by having the automated decision-aid.

<u>Usefulness/Utility</u>: The usefulness of data or machines means responding in a useful way to create something of value for operators, eventually developing into trust. One study by Morrison et al. (1998) examined the usefulness of decision support systems within the tactical decision-making environment. They provided several decision-making support modules that assisted human operators to understand the situations and makes judgments. Conducting an experiment with two operators as a team, they found that the operators tended to rate those parts of the decision-making support system modules that supported quick decision-making, and thus were more useful; higher than other modules.

This list seems comprehensive in that it includes factors beyond those of Muir (1994), or Rempel et al. (1985), that may have effect on operators' trust indirectly. For example, consider understandability and transparency. These factors may not be the ones that operators directly access when they evaluate the level of trustworthiness of an automated system. However, as they understand the automated system better, they can grasp knowledge about the pros and cons of the system and when to rely on the system.

Therefore, it seems that Sheridan's list of trust attributes provide comprehensive understanding of operators' trust and systematic characteristics of trust that may cause operator's trust to develop.

2.2.3.3 Research on Trust in Automated Control Systems

Based on the research framework to study human trust in human-machine interaction setting, several studies included experiments in a continuous processing domain.

Among the prior research, two studies, Muir & Moray (1996), and Lee & Moray (1992), made attempts to identify the role of operator's trust in automated systems. Further, these two studies showed interesting results on the relationship between the operator's trust in automated systems and their reliance on the system measured by the use of the system. Conventional wisdom implies that the use of or reliance on automation becomes higher as the human operators trust the system more. For example, Muir & Moray (1996) performed a process control simulation experiment and the results supported this perspective, showing a positive regression coefficient between operator's trust and the use of automation. They also concluded that as the level of operator's trust suffers their use of automated systems decreases. Lee & Moray (1992) conducted very similar experiments and also showed that the operators tended to use less of the automated system as the level of trust decreases.

First, having defined operators' trust in automation as a simple additive model of three characteristics, Muir & Moray (1996) conducted a series of experiments to investigate the validity of Muir's (1994) definition of operator's trust in automation and the relationship between the properties of the automation, trust, and human intervention. The main systematic factor, reliability of the automated system, was controlled by adding either constant or variable errors. These errors were generated and added to the operators' input, the actual process, and the automated system's output. Note that they differentiated between the errors originated from controlling activities and the errors in the display. The level of trust was subjectively approached through a questionnaire. They found that human operators were able to express their trust in automation quantitatively and also found that Rempel et al. (1985) three characteristics were representative to estimate operators' general level of trust in automation. In terms of Barber's (1983) taxonomy of trust, Muir & Moray (1996) stated that competence was the better single predictor of trust. Additionally, they found that an "honest" automated system was trusted more than the other two systems, which contained either constant or variable errors. An interesting finding was that the two "poor" systems were not different in terms of operators' trust in the system. In terms of the dynamics of trust development, they found that faith factor played a critical role in the early stage of interaction, then predictability followed by dependability. It is quite contradictory to what Rempel et al. (1985) results demonstrated. Rempel et al. (1985) showed that predictability was the key role in an early stage, then dependability, followed by faith.

The second experiment by Muir & Moray (1996) was performed to investigate the differential effect of consistent or variable error on operators' trust in automated systems,

and the relationship between the level of trust and monitoring activities. They found first that trust was greatly reduced by small control errors, but was increasingly less sensitive to larger control errors. Operators' use of automated system was also affected by the magnitude of the control error. They also found that participants engaged more on monitoring activities when they trusted the automated system less. This is consistent with what Sheridan & Hennessy (1984) suggested: operators would spend more time monitoring automated systems they would not trust.

Second, Lee & Moray (1992) conducted several similar experiments to investigate relationships between human operators' trust in automation and their use of the automated systems. Automation failures were manipulated in two dimensions: length and size. They controlled the length of automation, transient vs. chronic failure, to investigate the dynamic characteristic of trust. Also, this would reveal the length of recovery affected by different lengths of failures. Automation failures were also varied in size to see any differential effect on trust in automated systems. They showed a strong correlation between human operators' trust in automation and their use of the automatic controller, as well as an inverse relationship between operators' trust and their monitoring activities. Participants showed a decreased level of trust in automation when faults occurred in the automated system and engaged less in monitoring the automated system under the automatic mode. These results seem correct in that as the level of reliability decreases, human operators should trust the automated system less. They found out, first, that human operators were able to maintain their high level of overall performance despite the effects of a fault in the automatic controller. This is consistent with Muir and Moray's result. Participants were able to maintain a high level of performance while interacting with an unreliable automated system. In this case, the degree of reliability was defined as a percentage variation from the targeted values.

A behavioral counterpart of operators trust in an automated system is an index of reliance. This is based on the assumption that the more operators trust, the more they use. Riley (1996) investigated a variety of factors affecting the human use of automation simultaneously. He performed an experiment using a simple computer game in which participants had to control various simple tasks, which they could relinquish control to the automated system. Automation reliability (90 percent and 50 percent) was controlled by setting the probability that the automation would perform a task correctly. The automation faults were distributed over time and the experimental conditions, manipulated the level of workload and in the level of uncertainty. The level of uncertainty in a task was controlled by introducing abnormalities so that the automated system could not perform the task appropriately. The level of workload was also controlled by making another separate task more complicated. The results showed that participants demonstrated a bias toward manual control, and that both college students and pilots did not delay turning on automation after a failure and continued to rely on failed automation. Pilots showed greater tendency toward the use of automation under the lower risk. In a subsequent experiment focusing on the dynamics of trust, Riley controlled the level of information that participants were given about the automation prior to performing the experiment. These conditions differed in whether participants were provided with information regarding automation reliability, state uncertainty, or both to

reveal the contributions of each element to their automation use decisions. He found that both state uncertainty and trust affect automation use decisions, but only early in the participants' experience with automation. Based on a variety assumed relationships affecting operators' use of automation, he found several interesting relationships, shown in Figure 3.

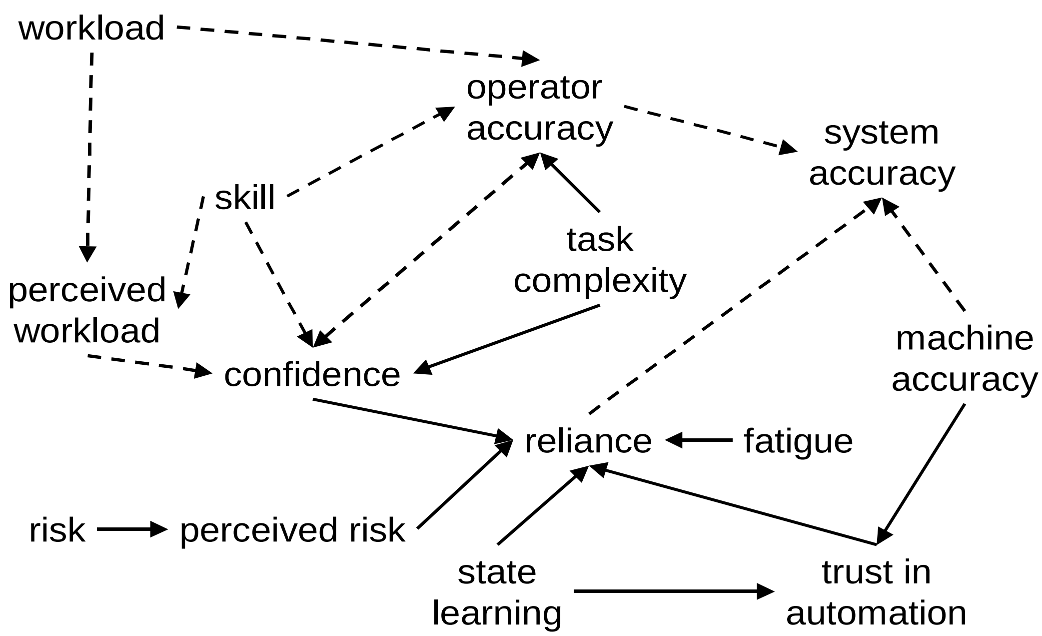

Figure 3: Web of Factors Affecting Operators' Use of Automation (Riley, 1996)

Additionally, Itoh & Tanaka (2000) developed a mathematical model of trust in automation in which a variety of combinations of the three characteristics proposed by Rempel et al. (1985) were used in the evaluation criteria to determine whether to rely on the automation or not. The major point that makes this study different from others is that this study considered the state of the automated systems in determining trustworthiness, thus addressing the concept of calibration of trust.

Although those research projects concentrated on human trust in automation based on the general understanding of supervisory control tasks, the importance of the trust concept seems applicable and attempted to define the role of trust in other domains, such as computer supported cooperative work (Jones & Marsh, 1997; Christianson & Harbison, 1997), decision-making in management (Lerch & Prietula, 1989), medical diagnosis expert systems (Moffa & Stokes, 1996), computer security problems (Beth, Vorcherding, & Klein, 1994), and trust in e-commerce (Gefen, 2000).

2.2.3.4 Research and Models of Trust in Decision-aids or Expert Systems

Specific empirical research on operators' trust in automated decision-aids is somewhat more limited than general research on trust in automated systems. Several studies used different types of decision-aids that might have some relevance with the issues in this study.

As discussed before, Glover et al. (1997) conducted an experiment investigating the relationship between decision-aids and humans. Among the results, they showed that the users' performance was significantly better with the decision-aid than without. This was only true for those typical cases. The reverse was the true for those cases that were not typical. The results meant that participants were able to perform the task better without the decision-aid for the cases in which the decision-aid was not designed and was not expected to encounter. Also, it indicates that participants relied on the decision-aid inappropriately for those cases that the decision-aid could not perform. In conclusion, decision-aids that are inferior for non-typical cases can cause operators to inappropriately rely on the automation, which consequently results in a low level of performance.

For instance, one study by Dzindolet, Pierce, Beck, & Dawe (1999) investigated different aspects of automated decision-aids on operators' identification performance and revealed that participants decided to ignore the decisions by the decision-aid when no feedback was provided, and when they were told that the decision-aid was actually human rather than an automated aid, regardless of the quality of the decision-aid (better than or equal to participants performance). However, the percentage of participants who decided to ignore the support from the decision-aid when they were told that the decision-aid was actually an automated decision-aid was also not promising, regardless of the presence of the feedback or the quality of the decisions. The average percentages of participants who decided to ignore the decision-aid were 83 percent and 81 percent for the feedback and no feedback condition, respectively.

Cohen, Parasuraman, & Freeman (1998) proposed a qualitative model of trust that allows for the operators' ability to distinguish the situations in which the automation should exhibit better or worse performance. They proposed the Argument-based Probability Trust (APT) model, which borrowed from terminology used in Toulmin's (1958) analysis of logic and argument construction. The main idea of Toulmin's model of argument consists of five components: warrant, grounds, qualified claim, backing, and rebuttals. The final outcome or conclusion can be supported by evidence or grounds that should necessarily be linked by warrant. A warrant refers to a specific kind of link or rule, i.e., if-then rule, that grounds satisfy the claim. Additionally, the existence of backing provides theoretical or empirical explanation of the warrant. Using the model of argument to interpret operators' trust in automation, they described trust in decision-aids based on event trees, which consists of nodes and branches. Branches represent possible alternatives of a factor of the environment that can affect operators' judgment and their trust while nodes represent the user's trust in the decision-aid at that point. The level of user's trust is calculated by assigning the user's belief of each branch.

Lerch & Prietula (1989) conducted an experiment consisting of traditional financial management decision problems to investigate the effect of inferior quality advice on people's judgments. As might be expected, there was a loss of trust in the agent resulting from faults in the advice, and the recovery of trust was slower than that in performance. They called this "inertia," similar to what others have called a "hysteresis loop." They also found that it was more difficult for humans to recover trust once it was lowered because of the failures to build trust initially. In Lerch & Prietula (1989), the level of performance measured by the level of confidence in the decisions the participants made deteriorated after the wrong advice and never returned to the level of performance where it was before the wrong advice, even at the end of trials.

Finally, Bisantz & Seong (2001) defined different types of failures that could occur in information acquisition and information analysis types of automation, acknowledging that systems could fail due to different intents. For instance, system components or automation may fail due to unintentional hardware or software difficulties. In some environments, failures may be introduced intentionally through acts of sabotage. They conducted an experiment investigating effects of different types of failure on operators' decision-making performance and trust in two informational sources. One informational source, the information window, provided sensed information by various sensors while the other, the decision-aid window, provided the environmental estimate of an object's identity. Participants were told that failures due to hardware or software failures or sabotage are possible. A third condition was also included, where the probability of decision-aid fault was not mentioned. Note that operators could select whether to consult with the automated decision-aid. Results showed significant impact of fault conditions on the information window use. Participants in the sabotage condition, seemed least likely to select the information window on a track-by-track basis, and participants in the control condition were more likely, across the three conditions, demonstrated a trend toward decreased information window use.

They concluded that there appeared to be a trend for participants in the sabotage condition to make less use of the information window than participants in the other two conditions. Also, participants showed a trend across conditions to make less use of the information window and more use of the decision-aid window, indicating that participants attributed the failure of the information window rather than the decision-aid window.

2.2.3.5 Measurement of Trust

While many studies (e.g., Larzelere & Huston, 1980) have justified the use of measurements using rating scales, it is also obvious that different measurement schemes, (e.g., percentage of time that operators used automatic controllers and frequency of monitoring activities) should be developed to predict the operators' actions in a human-machine interaction environment. Several studies of trust have been conducted from this perspective.

As discussed in the previous section, while studies on people's trust in an interpersonal relationship are rich in the theoretical perspective, little has been done on empirical measurement of people's trust.

Rempel et al. (1985) developed an 18-item questionnaire focusing on three characteristics of trust. An important question is how to validate whether a questionnaire measures a construct such as trust. This is less of an issue for human factors oriented studies of trust however, for which the construct of trust in automated systems can be associated with some qualitative behavioral measures, such as reliance on the automated system.

To date, there is only one study systematically examining the multi-dimensional characteristics of trust and identifying appropriate words or terms to describe people's trust in automated systems. Jian, Bisantz, & Drury (2000) performed an experiment in which participants compared and rated according their subjective relevance to the various characteristics of trust that they defined based on the sociological studies (Rempel et al. 1985). In addition to those characteristics defined in the previous studies, they found several words or phrases that could describe the term trust or lack of trust. The Jian, Bisantz, & Drury (2000) questionnaire study consists of twelve questions to which participants respond using Likert-type scales. Five questions are negative describing the automation's behavior that can cause a low level of trust and seven questions describe positive aspects of behavior. This questionnaire has been used in assessing the operators' trust in automated decision-aids (Bisantz & Seong, 2001). Bisantz & Seong (2001) found that the questionnaire based on the fault condition was sensitive to the difference in task conditions, and also that the different levels of fault conditions were affecting participants' subjective responses of trust differently. Kim & Moon (1998) performed a similar study investigating trustworthiness of cyber-banking system interface, and found that the words that are closely correlated with system's trustworthiness are dependable, reliable, realistic, and uniform.

2.2.3.6 Summary

Research on people's trust in interpersonal relationships reveals that trust is not a simple characteristic that can be easily assessed. Rather, it has multiple facets, each of which may require different evidence to grow and which may require a substantial period of time to reach a level where individuals can truly trust the other partner (Rempel et al., 1985). The characteristics identified provided us with a better understanding of the concept of trust in interpersonal relationship which can be extended to the human machine interaction perspective.

Literature seemed to suggest that the term within the human automation interaction domain also has multiple characteristics associated with it. As a definitive study, Sheridan's (1988) characteristics of trust in automated systems are described. Empirical research and models of trust in automated control systems are described to provide a better understanding of the studies on operators' trust in automated systems. From these studies, it seems there is a common consensus between these studies that use

of automated systems is tied to the level of trust. Finally, existing methods of measuring trust were described.

2.2.4 Summary of and Discussions on Background Research

In the previous sections, three major issues within the human machine interaction paradigm have been discussed to raise important research questions regarding human operators' judgment behavior and trust in conjunction with automated decision-aids.

First, a discussion on automated systems is provided describing how different types or levels of automation may stimulate or emphasize different aspects of operators' trust in automated systems, which consequently may have a different impact on the operators' use of the automated systems. In this perspective, human centered automation was discussed to provide deeper understanding of the types and levels of functions that human operators interact with the decision-aids. Studies regarding the taxonomy (e.g., types and levels) of automation seem to indicate that a variety of combinations of different levels and types of automation are possible depending on a variety of factors that can have impact on the decisions on the automation. Among these various alternatives of automation, automated decision-aids are characterized as the systems that every type of automation but the decision selection is designed at a high level of automation. Finally, the automated decision-aid in the current research was defined and described to outline differences from the definitions provided in the previous studies on the types or levels of automation.

Several studies were described to introduce different aspects of automated decision-aids that may have an impact on the operators' judgment performance or the use of such systems. There seems to be a positive relationship between "imperfect" automated decision-aids and operators' performance, which is somewhat consistent with the results from the previous studies on operators' performance with automatic controllers. However, those decision-aids used were designed to support the human operator's problem solving activities. Therefore, it is difficult to identify whether there was any differential effect of the different magnitudes of failures on operators' performance. In other words, we need to first be able to evaluate the competence of the automated decision-aids, and then it may allow us to understand the differential effect on operators' judgment and decision-making performance. One important issue regarding the relationship between operators' judgment behavior and performance has been identified in this process: operators' trust in automated decision-aids.

Second, research on trust was described to establish the concept and role of trust within this research framework. First, trust in interpersonal relationship was described. Established from the sociological perspective on trust in interpersonal relationships, discussions on human operators' trust in automation together with justifications on the importance of this study on operator's trust in automated decision- aids has been provided. Briefly, two important studies (Lee & Moray, 1992; Muir, 1994) revealed that operators' trust in automated control systems plays a critical role in determining the use of automated control systems. Discussions on two streams of studies on trust in automation in general were provided: trust in automatic controllers, and trust in decision-

aids. Rather than using automated controlling systems, several studies used the decision-aids in investigating operators' knowledge acquisition (Glover et al., 1997), decision confidence (Will, 1991), self-serving or automation bias (Dzindolet et al., 1999), and trust in expert systems in financial domain (Lerch & Prietula, 1989). The main difference between these two types of automated systems lies on the methods of engagement with the automated systems. This is important because it can be one of the objective measures to interpret operators' trust in such systems. First, with the automatic controller, the human operators have to make conscious decisions to engage in an automatic mode of control at least to understand the capability of automation, or how well the automated system performs the tasks compared to what it is suppose to do. On the other hand, with automated decision-aids, operators do not have to rely on the estimates to understand various characteristics of the automated decision-aids. Rather, operators have more chances to evaluate different aspects of the system, such as its capability, competence, or reliability, without actually engaging it. This can allow us to understand the relationship between the operators' judgment performance, various characteristics of the automated decision-aids, and trust in such systems.

Finally, the issue of measuring trust in automated systems was discussed to introduce existing methods both qualitatively and quantitatively. Having reviewed important issues in human machine interaction, specifically automated systems, let us turn our attention to constructing a general model of human operators' judgment and trust in conjunction with automated decision-aids.

2.3 Human Trust in Virtual Teams

One of the main purposes of this study is to examine potential multiple characteristics of human trust in virtual teams deduced from the variety of research conducted in many environments on the issue of human trust in general. There is limited literature with regards to the human trust in virtual teams. To make matters worse, there is no empirical research conducted to examine the role and importance of human trust in virtual teams, although the practical implications are very obvious. Therefore, some studies on this topic are reviewed and summarized (see Appendix A for the list of literature reviewed in this study).

Trust development in virtual teams presents significant challenges because it is difficult to assess teammates' trustworthiness without ever having met them. Moreover, as the life of many virtual teams is relatively limited, trust must develop quickly (Jarvenpaa & Leidner, 1999). Yet, trust development is deemed crucial for the successful completion of virtual team projects (Sarker, Lau, & Sahay, 2001).

Early work on trust in the virtual environment has found that short-lived teams are in fact able to develop high levels of trust but they do so by following a swift trust model rather than the traditional model of trust development (Jarvenpaa, Knoll, & Leider et al., 1998; Jarvenpaa & Leidner, 1999). The swift trust paradigm suggest that, when they don't have enough time to slowly build trust, team members assume that others are trustworthy and begin working as if trust were already in place while seeking confirming or disconfirming evidence throughout the duration of the project (Meyerson, Weick, &

Kramer, 1996). Virtual teams that exhibit high trusting behaviors experience significant social communication patterns, substantial feedback, positive leadership, enthusiasm, and the ability to cope with technical uncertainty (Jarvenpaa & Leidner, 1999). This early research identified perceived integrity of other team members as particularly important in the development of trust early in a team's life and perceptions of other members' benevolence as a trait that supported the maintenance of trust over time (Jarvenpaa et al., 1998). High trust teams may also develop as a result of early face-to-face meetings with the intent of developing a strong foundation of trust between members.

Gallivan (2001) summarized the types of trust that have been researched in the organizational settings. Those are:

- *Knowledge–based trust*: trust based upon a prior history of transactions between two parties.
- *Characteristic–based trust*: trust that is assumed based on certain attributes of the other party.
- *Justice–based trust*: a trusting environment, as ensured by guarantor agencies.
- *Swift trust*: a fragile form of trust that emerges quickly in virtual workgroups and teams.

Similarly, Sarker, Valacich, & Sarker (2003) defined three different bases of trust applicable to virtual teams. Those are:

- *Personality–based trust*: Researchers have shown that trust develops during infancy when one seeks and receives help from one's caretakers (Bowlby, 1982), and for many individuals, this results in a general propensity to trust others (Rotter, 1967). De Vries (1999) proposes that personality–based trust has a lasting effect and can be seen in individual's everyday interaction with others.
- *Institutional–based trust*: Drawing on institutional theory, this trust holds that norms and rules of institutions such as organizations surrounding individuals guide their behavior. Coutu (1998) refers to this trust as the deterrence–based trust, where team members will trust simply because of fear, that if they do not, they will be punished.
- *Cognitive trust*: Lewis & Weigert (1985) stated that "we cognitively choose whom we will trust in which respects and under what circumstances, and we based the choice on what we take to be 'good reasons,' constituting evidence of trustworthiness." As individuals get to know others, they gain more information about them. This information is processed through a sequence of stages in their minds, such as attention, recognition, evaluation, categorization, organization and storage retrieval, integration, and judgment (Feldman, 1981).

Although most of these forms of trust have been widely discussed in the literature, the last term, swift trust, was coined recently. Except this term, this categorization of trust is precisely the point of theory of trust developed in this study which will be described in detail later. First, a review of trust in the current literature is provided to set the ground for the development of a robust theory of trust.

Bulu & Yildirim (2008) investigated (preservice teachers') trust levels and collaborative communication behaviors namely leadership, feedback, social interaction, enthusiasm, task and technical uncertainties, and task-oriented interactions in online learning environment. This is based on others' premises (Handy, 1995) that "trust needs touch." However, the results from Bulu & Yildirim (2008) indicated that although the groups developed their work in an online environment, some of the groups ended up with higher group trust levels. *This showed that trust could be built in online groups.* As similarly discussed by other studies (Ishaya & Macaulay, 1999; Jarvenpaa & Leidner, 1999), their study indicated that the types of interaction or the types of communication behaviors, including distribution of the communication, feedback, social interaction, enthusiasm, coping with technical problems, and task oriented interactions may affect trust levels of the online teams. Also, it showed that the groups with different trust levels showed different communication behaviors throughout the process. This indicates that certain collaborative communications behaviors should be presented by the team members for successful collaborations, and stronger trust levels in online teams. This result is similar to Coppola, Hiltz, & Rotter (2001) showing that predictable communications foster the later trust development. They also found that initial enthusiasm and social interaction can help build trust in the early period. However, they noted that it should be continuous to maintain trust. As Coutu (1998) succinctly summarized, teams with the highest levels of trust tended to share three traits. First, they began their interactions with a series of social messages – introducing themselves and providing some personal background – before focusing on the work at hand. This initial period of electronic "courtship," appears to be particularly important in establishing *knowledge–based trust*. Second, they set clear roles for each team member. Assigning each member a particular task enabled all of them to identify with one another, forging a foundation for *identification–based trust*. The third hallmark of the trusting team had to do with attitude: team members consistently displayed eagerness, enthusiasm, and an intense action orientation in all their messages. "One pessimist has the potential to undermine an entire virtual team," (Jarvenpaa, Knoll, & Leidner, 1998).

Jarvenpaa, Knoll, & Leidner (1998) examined how trust develops in virtual teams. Teams of university students worked on computer–mediated problem–solving simulations. Findings indicated that trust was positively related to virtual team productivity. Trust in teammates' task competence developed in the early stages of the simulation whereas trust in teammates' benevolence developed later (their factors were trust, trustworthiness, ability, integrity, benevolence, and propensity to trust – the last four were referred as the antecedents of trust). The authors concluded that trust is 'pivotal in preventing geographical distance from leading to psychological distance in a global team.' In the context of the computer simulation, they also experimented with team–building activities before the start of the project. Surprisingly, these activities did not have a direct effect on trust levels. This raises some interesting questions on the relationship between trust and virtual team performance and the most effective strategies for building trust within virtual teams. Yet to be investigated are the specific behaviors of virtual team members that contribute to perceptions of trust or mistrust. Figure 4 shows

their model based on the experimental results. The questionnaire "trust measures" used in this study are shown in Appendix D.

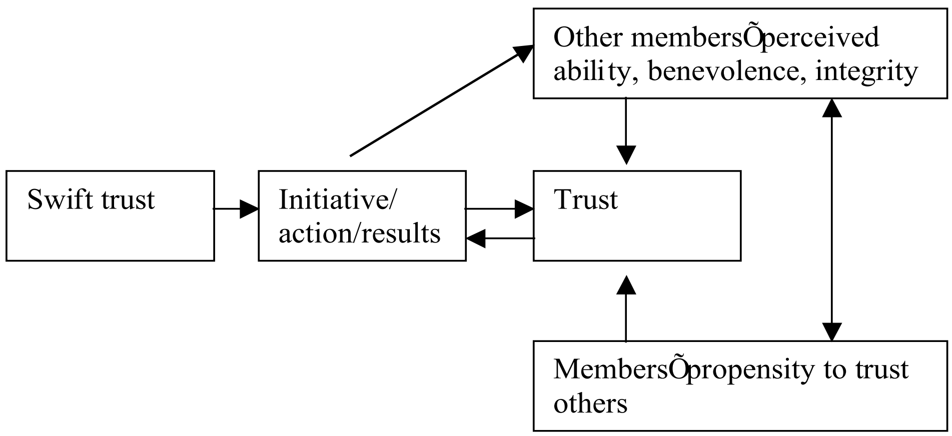

Figure 4. Relationships among Factors Affecting Trust from Jarvenpaa, Knoll, & Leidner (1998)

Thomas & Bostrom (2008) investigated the effect of leadership style (Theory X vs. Y) and found that Theory X (command and control) and Theory Y (facilitate and support) leadership approaches enable and contradict effective technology adaptation, trust building, and outcomes. Successful adaptation led to increased trust. As shown in Figure 5, changes in trust and cooperation as well as technology adaptations were associated with better team outcomes. This is an interesting finding, as the authors pointed out, because it supports the notion that any leader in a virtual team should have at least some basic awareness and skills for managing technology adaptation. Although there has been much speculation over the years on the nature of leadership style in achieving outcomes in virtual and knowledge work settings, which indicates that a more facilitative Theory Y orientation would generally be more important and effective, especially when it comes to forming trust. Results from Thomas & Bostrom (2008) showed an important role for Theory X style leadership involving mandates, controls, and command as these actions were significantly and directly related to improved outcomes, as well as, improvements in trust and cooperation. The results also suggested that moiré facilitative and supportive Theory Y actions were not more important for affecting trust and cooperative directly. This model is shown in Figure 5.

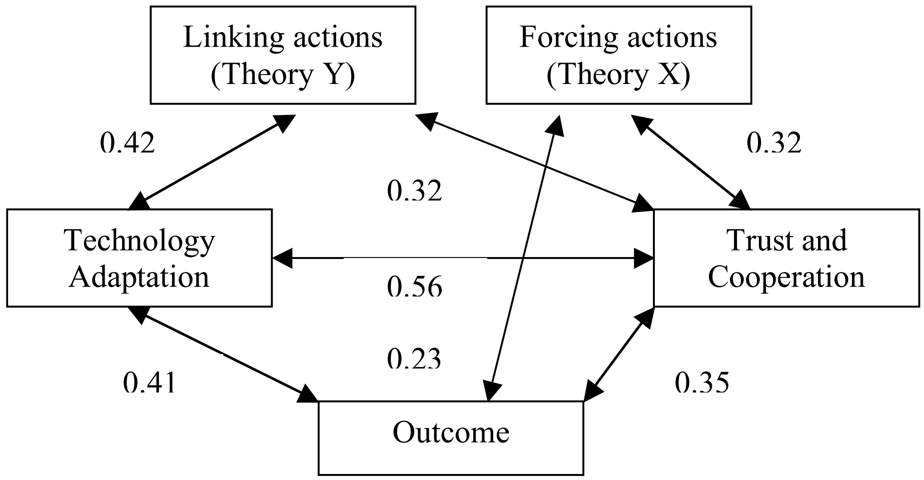

Figure 5: Correlations Mapped to Research Model Used in Thomas & Bostrom (2008)

Additionally, human trust has been characterized as an internal construct that is developed over time. Regardless of the settings that the studies investigated this dynamic characteristic of human trust, interpersonal or human machine interaction environments, many studies showed and supported this notion. However, according to Iacono & Weiband (1997), Meyerson et al. (1996) argued that a different form of trust, swift trust, is necessary for temporary teams. Situations requiring swift trust do not allow for the incremental, gradual, evolution based on increasing knowledge and information about others; behaviors. Since the temporary group must move forward quickly to accomplish goals, members must have act swiftly as if trust were in place rather than waiting to see who can be trusted and who cannot. They must work continuously and consistently to maintain expectations of trust.

Attempting to find relationships between performance and trust in business and organization environments, Iacono & Weiband (1997) found that:

1) high performing teams quickly began to form teams,
2) high performing teams did not get bogged down in evaluating and responding to every initiation, and
3) high performing teams were able to handle several activities at once.

Additionally, there were six categories that they used to measure trust; getting together, work-process, work-content, work-technical, needing contact, and fun talk. They found that work-process initiations and responses were positively correlated with team performance. Similarly, work-content initiation and responses were strongly related to how well the team performed on the work. Also, they found that lack of continuous interaction and the struggle to meet deadlines at the last minute contributed to disconfirming evidence in the reliability of team trust. One of the interesting results from this study is that unlike the expectations of swift trust, distributed teams did not focus exclusively on work-content or on activities that can be characterized as "staying on task." Their results indicated that in distributed, temporary teams where solutions about

who does what are always emergent and negotiable, high levels of work process interaction would be required.

Another interesting study with regards to the dynamic nature of trust, Kanawattanachai & Yoo, 2002), empirically examine trust and the differences between high– and low–performing virtual teams in the changing patterns in cognition– and affect– based trust over time. In this study, interpersonal trust was recognized as a multidimensional construct with both cognitive and affective foundations.

Cognition–based trust refers to the calculative and rational characteristics demonstrated by trustees. These include reliability (Rempel, et al., 1985), integrity, competence, and responsibility. People assess trust based on various attributes such as certain types of professions and levels of familiarity. As the authors stated, one basic function of cognition–based trust is to reduce the complexity among social actors. Trust is strengthened when trustees do what they promise to do in a timely and professional fashion. On the other hand, *affect–based trust* involves the emotional elements and social skills of trustees. Care and concern for the welfare of partners form the basis for affect–based trust (Rempel, et al., 1985). Unlike cognition–based trust, affect–based trust has typically been studied in the context of close social relationships such as couples, family members and friends. Although affect–based trust is typically found to be important in the context of close social relationship, McAllister (1995) found that even in working group environments, it influences the performance and well-being of the teams in such a way that one takes another's problem as his own and is willing to give a helping hand to the needed party even if they did not request assistance.

Experimenting with MBA students across several universities competing in a web-based business simulation game over an 8 week period, they found that both high– and low–performing teams started with similar levels of trust in both cognitive and affective dimensions. However, high-performing teams were better at developing and maintaining the trust level throughout the project life. Specifically, the results showed that in virtual team environments, the level of cognition-based trust was higher than that of affect-based trust throughout the course of the project, regardless of the team performance. Also, they attempted to identify the differences between high– and low–performing virtual teams in terms of cognition– and affect– based trust at the outset of the project. They found that there were significant differences between the two teams in terms of the cognition-based trust at the early period. This is the evidence that a large number of high-performing teams in high-trust teams, which reflects the presence of swift trust, and a large number of low-performing teams to be found in low-trust teams, which reflects the absence of swift trust at the later stage of interaction. ***Moreover, virtual teams relied more on a cognitive than an affective element of trust***.

In addition, trust has been studied combined with many other factors, such as user satisfaction (Morris, Marshall, & Rainer, 2002). Finally, readers are directed to Powell, Piccoli, & Ives (2004) for the summary of current research with regards to virtual teams. The summary table from their paper is shown in Appendix E and Appendix F contains a summary of major issues investigated in these studies.

Based on this literature on human trust in virtual teams, a theory of trust was developed as a tool to provide deeper understanding of how human trust can be assessable and should be researched. The theory of trust stems from Hammond and his colleagues' work (1987) on the cognitive continuum theory. Therefore, this theory is reviewed in the next section to set the foreground for the theory of trust.

2.4 Cognitive Continuum Theory

Cognitive continuum theory with the social judgment theory and the Lens Model representation showed some specific effects, predictable from this theory, of task manipulations on human decision-makers' location of the cognitive continuum, measured in the magnitude of the Cognitive Continuum Index (CCI). This index was formed based on four cognitive properties obtained from the Lens Model analysis: cognitive control, organizing principles, error distribution, and differential confidence in process vs. outcome.

According to Hammond, Hamm, Grassia, & Pearson (1987), the intuitive end of the continuum is characterized as rapid and simultaneous information processing, not easily traceable judgments (if ever), not available formal logical reasoning, high confidence in judgment outcomes and low cognitive demand, while the analytic end of the continuum indicates the opposite. More importantly, based on Brunswik's (1952) definition, quasi-rationality between the two extreme poles was characterized as involving both aspects and being used in this region in our daily life. Depending on the demand or configuration of the information processing task environment, this may be used more or less intuitive or analytical.

Among the properties suggested by Hammond et al. (1987), some are worthy of mention due to their potential implication for the issue of design for trust. Such characteristics include reliance on pictorial cues or non-verbal cues, rapid information processing, and processing of simultaneous cues evaluated at perceptual level rather than "measurement" level. These characteristics impose a few requirements on designing a graphical interface to induce an intuitive mode of cognition. Similarly, the characteristics suggested for an analytic mode of cognition requires that information needs to present in a quantitative way in order to allow human decision-makers to engage in high levels of cognitive control, or high confidence in the process of judgment and decision-making process and performance.

Additionally, one such characteristic, responsiveness to new cues, has a major implication for a decision-making framework. Specifically, in an environment where judgments must be made based on a sensor or a group of sensors, and the network is subject to any malicious exploitation, in the typical tactical decision-making framework, human operators' decision-making process must be induced to engage in an analytic mode of cognition so that the operators can take the novel cues or situation into consideration.

2.5 Theory of Trust: Trust Continuum, Trust Grid, and Trust Cube

Unlike the traditional classification of trust being either high or low considering its dynamic characteristic, it seems the notion of a continuum used by Hammond and his colleagues can be a better fit to describe human trust. In this section, the theory of trust, trust continuum, trust grid, and trust cube are described.

2.5.1 Trust Continuum, Trust Grid, and Trust Cube

Borrowing the cognitive continuum theory (CCT) and its index (CCI) to define the continuous mode of cognition, it is proposed that a trust continuum can be defined as a continuum whose end poles are similar to those in the cognitive continuum; intuitive and analytic (shown in Fig. 6). However, it is somewhat difficult to understand what "intuitive (or analytic) trust" could be within the continuum. Among the characteristics used by other studies including Sheridan's seven characteristics, trust continuum can be mapped and superimposed to describe how each of these characteristics can be placed as part of the human operators' trust in automated systems.

Trust characteristics	Trust Continuum
Explication of Intent	Analytic
Reliability	
Competence	
Dependability	Quasi-rational
Usefulness/Utility	
Understandability	
Familiarity	Intuitive

Figure 6: Trust Continuum Mapped Out Sheridan's Seven Characteristics of Human Trust in Automated Systems

Analytic trust refers to an aspect of trust that may be empirically established through interaction among humans and automated systems. Sample metrics include reliability and robustness. However, intuitive trust must be derived subjectively. Moreover, sample metrics include issues such as bias, faith, and experience that are prevalent in human interactions with automation as demonstrated by graphics and displays. The observed subjective data may be converted to measure such metrics as understandability and utility. These features of the analytic-intuitive trust continuum are similar to those features in Hammond et al. (1987) definitions of intuitive and analytic cognition.

Rooted from the Hammond's cognitive continuum theory, the Sheridan's seven characteristics of trust can be "intuitively" categorized and placed along the scale of the trust continuum, varying from intuitive to analytical. This categorization is based on the types of communication method/channel necessary to educate the human counterpart of the automation's characteristics.

From analytic to intuitive aspects of trust, the order of these characteristics are; explication of intent, reliability, competence, dependability, usefulness, understandability, and finally familiarity. Therefore, explication of intent is placed as the most "analytic" characteristic while familiarity is the most "intuitive" aspect of trust.

More importantly, human trust has been known to be empirically very fragile and difficult to recover once broken regardless of whether it is looked at from a sociological or human-machine interaction perspective. Therefore, it is important to provide any training implication or graphical interface, if necessary, so that human operators can calibrate or "recover" their trust depending on the trustworthiness of the automated systems. One attempt (Seong & Bisantz, 2008), using the idea of cognitive feedback based meta-information showed somewhat positive results in that those provided with meta-information feedback were able to calibrate their trust according to the automated decision-aid's trustworthiness, which was evaluated based on the competence and the notion of how reliably the system can generate estimates for the unknown objects. In this study, the meta-information feedback is based on Sheridan's notion of understandability; how the automated decision-aid integrated the sensed information to produce the environmental estimates. Once dealing with faulty automated systems and provided with the meta-information feedback (note that this is classified as closer to the intuitive end of the continuum, as shown in Fig. 7), participants were able to utilize this information to further calibrate their trust. Note that the cues or information is not new, rather participants were given faulty estimates.

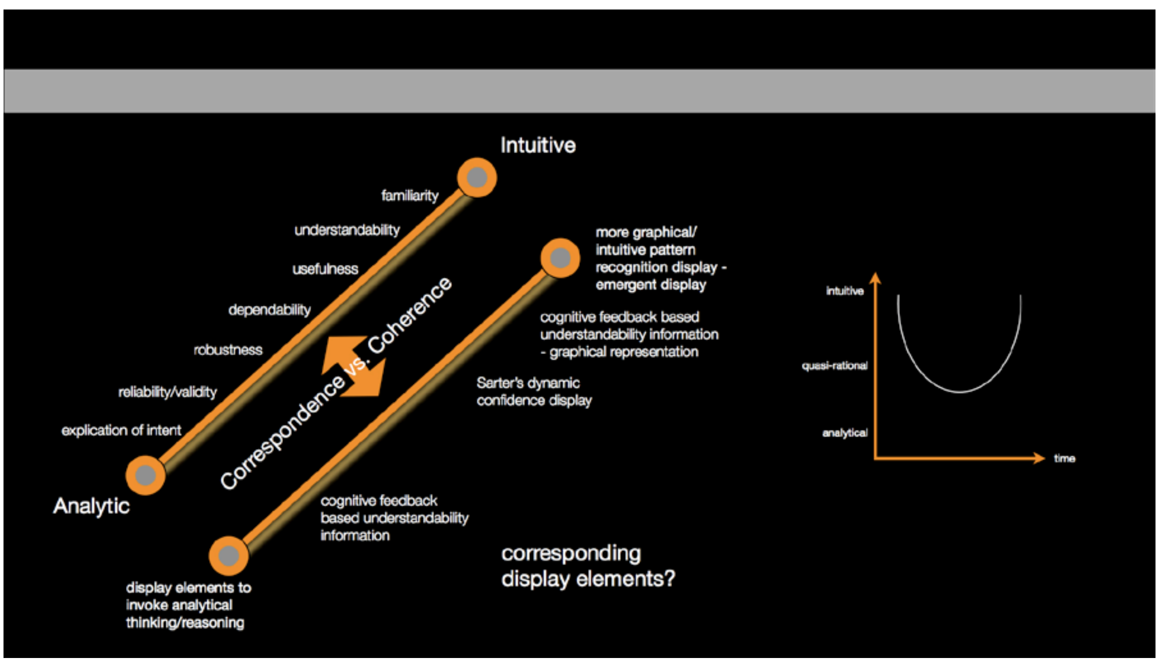

Figure 7: Trust Continuum Along with Sheridan's Seven Characteristics and Corresponding Potential Practical Systematic Components to Support Human Trust Calibration. The Figure to the Right Shows the Transition of Cognitive Mode (adapted from Hamm, 1988)

Theoretical approaches in analyzing human trust in automated system have been explored in the above section. In summary, trust is an internal state of human operators, but a state with proven links to antecedents within the operator and in the task and proven effects on strategy, performance and perhaps operator well-being. Because of this structure, trust has the same similarities to stress. Stress is an internal construct based upon the operator's perception of themselves and their task, both perceptions being influenced by, but not identical to, objective reality. Similarly, stress can drive actions of the operator, in terms of how to cope with the task. Finally, stress cannot be measured only with reference to externally – verifiable phenomena: the operator's perception is a required ingredient.

Not having a direct measurement for the construct of trust due to its' internal characteristic gives us another difficulty to understand and measure it. Therefore, a measurement for such an internal construct as trust should be validated regardless of the methods utilized to measure it. Campbell and Fiske (1959) proposed an analysis technique called Multitrait-Multimethods Matrix (MTMM), to determine whether experimenters are measuring what they want to measure rather than something completely different. While convincingly demonstrating the faults of single-concept, single-operation methodologies in the field of individual differences, they showed that most results could be determined and are dependent upon the methods. More recently, the method was picked up by Hammond and his colleagues (1987) to apply to the judgmental analysis framework. As they pointed out, the technique has not been used in any recent experimental psychology. Also as pointed out by Vicente and his colleagues, the technique is rarely used in the human factors domain. One of the studies that used the MTMM is Lee's study. In this research, he measured two constructs, trust and self-confidence, with two different methods. In this way, the validity of whether such methods used (subjective scales and monitoring behavior) could be argued to effectively measure the two constructs. He found that there is a very low correlation between trust, which is measured by a subjective scale and by monitoring behavior. By this, it could be argued that the experiment was not valid in that the methods did not measure the constructs properly. This is what Campbell and Fiske (1959) called the lack of "convergent validity," which explains different methods to measure the same construct should yield same result. Further, this might imply, that the participants did not treat the subjective scale of trust the same as on the automated system, therefore they answered in different ways. This might be another implication of biases (either overconfident or less confident than what they should).

Having defined the term trust as a multi-faceted construct, here it may be suggested that by measuring each characteristic instead of measuring trust in total, the major factors among the defined characteristics could be identified using this technique. Therefore, identifying the most effective characteristics could lead us to design or induce the more harmonized balance or calibration of trust according the true level of trustworthiness of the automation so that automation could be used somewhat more effectively. Human trust in automated systems has been acknowledged as a major factor in determining the degree of effective use of the systems. However, other variables have been suggested to have effects on the automation use.

The application of the MTMM matrix imposes another barrier to firmly understand the construct of trust. In the MTMM matrix, each trait to be measured should be by different methods. As stated above, the defined characteristics are very qualitative which do not offer any other measurement technique.

Understanding of the analytic-intuitive trust continuum must be examined via an actual human perception of trust in automation. This perception may occur at levels on a continuous scale from low to high. For simplicity of measurement, a discrete binary metric of low trust or high trust is used in this study. Thus, the interaction of the two factors on the trust continuum and on the two levels of trust perception yields a two dimensional metric of the trust continuum and trust level that leads to the development of four quadrants on a trust grid as demonstrated in Figure 8.

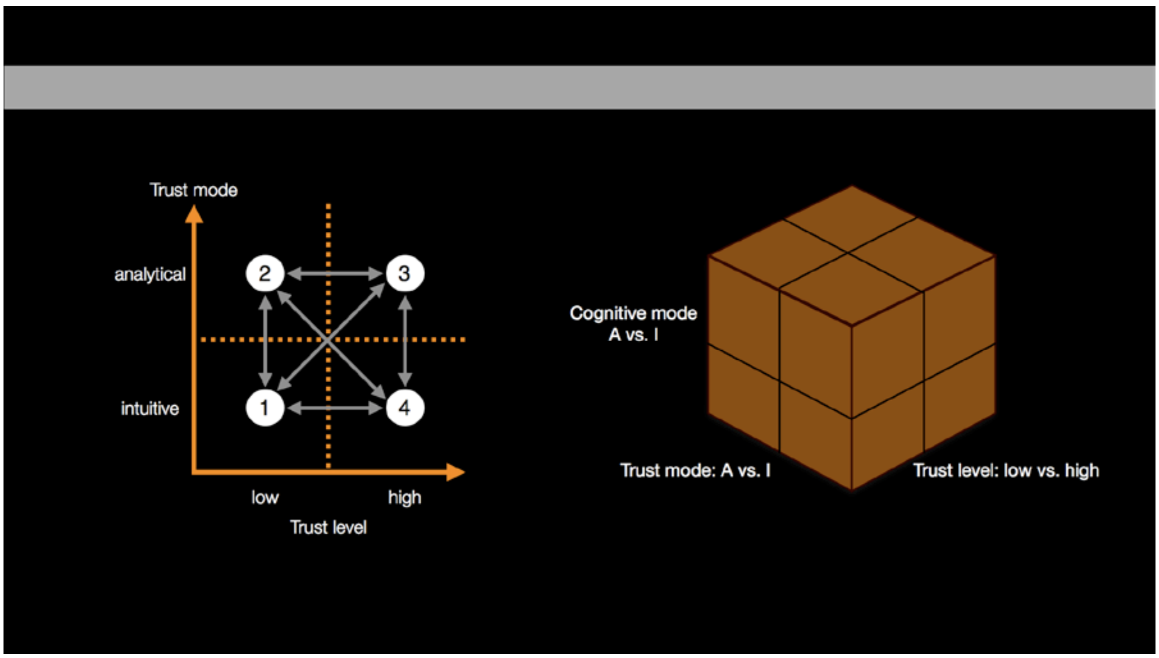

Figure 8: Trust Grid (Left), and Trust Cube (Right)

These four quadrants indicate that:

(I) **Intuitively derived trust and low trust level**: In this situation, trust is based on psycho-social factors such as bias and faith, and the interaction of these factors with poor automation leads to the perception of low levels of trust.

(II) **Analytically derived trust and low trust level**: In this situation, the human operator empirically combines trust factors to derive a single composite score of automation performance that allows for his/her perceptual evaluation of low trust in automation. This result may be attributed to many physical factors such as automation reliability (failure rate), human-induced errors, robustness, and consistency among human and automation decision-making.

(III) **Analytically derived trust and high trust level**: This quadrant captures the situation where human trust has been developed through consistent interaction with automation that demonstrated "good" behaviors. Contradictory to results from previous studies, which suggested that trust was fragile, analytical/high trust, may not be inherently fragile if humans understand a system's inner workings. This finding suggests that a human operator can precisely locate faults and adjust their decision-making strategy as appropriate for the situation.

(IV) **Intuitively derived trust and high trust level:** In this situation, humans overuse and/or over-rate automation through a high level perception of trust assignment. Intuitively, this result may be a result of positive bias and faith in automation. This quadrant captures the psycho-social aspects of human interaction with automation—imposing levels of expertise, experience, and training on the realism of automation behaviors. Both vulnerability and fragility of automation faults and human feelings may be captured in this quadrant.

Unlike the results from Hamm (1988), which indicated that human cognition started at the intuitive end of the continuum, gradually moved to the middle (quasi-rational), and finally moved back to the intuitive mode, it is the premise of this study that both ends of the trust continuum need to be examined concurrently in order to understand how each of the quadrants in Figure 8 affects interpretations of human trust in interactions with automated systems. The objectives and the experiments of this study are based on this premise.

Finally, this line of thought is designed to provide a proof-of-concept trust calibration using a two-interaction between two levels of trust and two levels of trust continuum. Beyond this study, the model can be extended to a multidimensional level in which the trust mode can be a factor—thereby making the model a three-dimensional trust cube (shown in Fig. 8). With this extension, it will be possible to theorize how trust can be measured along single and multiple slices of interaction of the three factors. The findings of this study will thus provide a better way to form theories regarding human-robot interaction systems (e.g., designing display and information presentations among humans and robots, determining when and how to enact feedbacks among agents, and determining how human factors and the robot behaviors interact to promote team task execution).

2.6 General Model of Automated Decision-aids, Trust, and Decision-making Performance

The automated decision-aid in this current study is defined as the "Decision-making" level within Endsley and Kaber's (1999) taxonomy of automation. This level of automation results in an interesting relationship between the environment, the automated decision-aid, and the human operators. The three components in Figure 9 illustrate the relationships which exist when human operators are assisted with an automated decision-aid in order to make judgments about the state of the environment. The three links shown in the figure represent opportunities for comparison between the components. The first link to be discussed is between the state of the environment and the human operators' judgment: "Human Operator Judgment Performance." Human operators assess the state

of the environment based on information provided, which may have an uncertain or probabilistic relationship with the state of the environment. This link represents the match between the operators' judgments and the actual environmental state, and therefore represents the human operators' performance in making judgments.

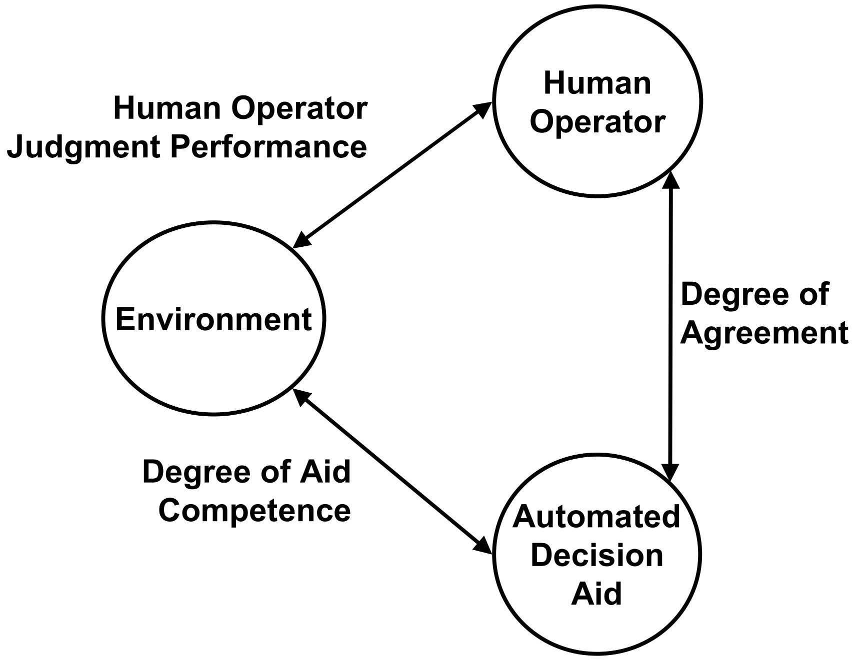

Figure 9: Three Components in this Study

Similarly, a link exists between the environment and the automated decision-aid. The automated decision-aid's task is to estimate the state of the environment based on available information, utilizing algorithms and techniques such as data fusion. A discrepancy is possible in this process when the automated decision-aid is not designed to anticipate the possible states of the environment or any malfunctions in processing the collected information occur. That is, the automated decision-aid may not be completely competent in estimating the state of the environment, which consequently may have an effect on the human operators' judgment performance when the environmental estimates are provided to them. One factor affecting the discrepancy between the true state of the environment and the estimates from the automated decision-aid is how well the uncertain or probabilistic relationship between the environment and the information available to the decision-aid is captured by and integrated into the automated decision-aid. Therefore, the discrepancy that may exist within this link can be used to evaluate the competence of the automated decision-aid; so labeled "Degree of Aid Competence."

The last link exists between the human operators and the automated decision-aid. This link represents the degree of correspondence between the judgments from both the components, denoted as "Degree of Agreement." Due in part to the fact that the information provided to either human operators or the automated decision-aid has an

uncertain or probabilistic relationship to the state of the environment, and also due in part to the limit of the human's information processing capability, there may exist a discrepancy between these judgments. This discrepancy can be either very small, in which case the judgments of the human operators and the automated decision-aid are mutually confirming in a certain degree, or very large, in which case the human operators must determine whether or not to integrate the automated decision-aid's estimate into the ultimate judgment. In these situations, the human operator's trust in an automated decision-aid becomes an important factor affecting the level of human operator's reliance on the automated decision-aid, which consequently may affect the overall level of the human operator's judgment performance. A contributing factor to this operator's level of trust is the degree to which the automated decision-aid provides information that is reliable, consistent, and valid. In those situations where information does not meet these requirements, the potential decrease in human operators' trust in automated decision-aid becomes a critical factor.

Figure 10 shows the flow of information among the system components and activities taken by the human operator and the potential role of human operator's trust in the automated decision-aid as a mitigating factor in determining the degree to which estimates from the automated decision-aids are utilized by the human operators. Measured or sensed information can be provided to both the human operator and the automated decision-aid so that each component can produce its own estimate. The estimates from the automated decision-aid are provided to the human operator so that the operators may incorporate those estimates to a greater or lesser extent depending on the quality of the estimates, and how much the human operator puts his trust in the estimates. This is represented in the figure by placing the human operator's trust in automated decision- aid between the automated decision-aid and the human operator. Although it has no effect on the quality of the automated decision-aid's estimates, the human operator's trust in the decision-aid may be a mitigating factor in the manner, which estimates from the automated decision-aid are incorporated into the final operator's judgment. It is possible that when the human operators do not trust the decision-aid because the decision-aid has shown either unreliable or invalid estimates, they may not take the estimates from the decision-aid into consideration to the final judgment. Additionally, the operators may fail to trust, and subsequently fail to use an aid because they do not understand the decision-aid.

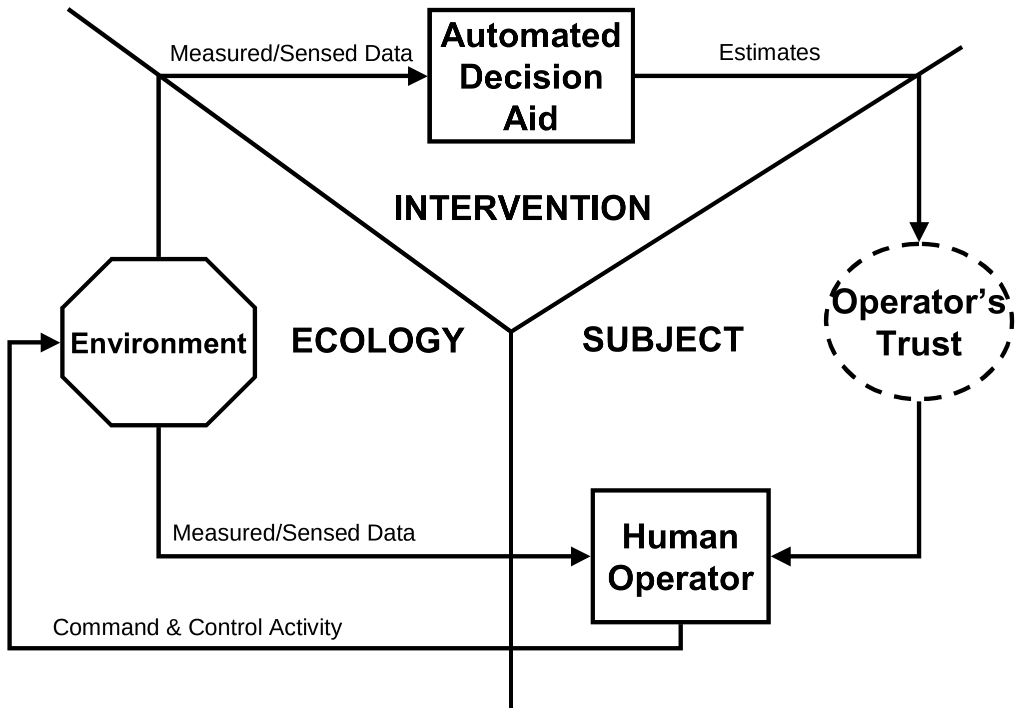

Figure 10: Flow of Information Among the Three Components in the System and the Role of Human Operators' Trust in Automated Decision-aid in Judgment and Decision-making

Important questions can be raised regarding the appropriateness of operators' trust in the automated decision-aid. As discussed above, human trust has been defined from various perspectives, such as sociological and human-machine interactions. Based on this literature review on the definition of trust, several system characteristics have been proposed which may influence human operators' trust in automated decision-aids including aid reliability, validity, utility, and understandability. These factors should normatively influence trust in, and reliance on automated decision-aids.

2.7 General Research Questions

Several classes of research questions were examined as part of this study. Although not extensive, there have been several studies on human operator's trust in automation. Based on the review of literature and conclusions drawn from the review, several research questions can be raised.

First, system characteristics of trust have been proposed in many studies from either sociological or human machine interaction setting, such as predictability, dependability, validity, and reliability. However, none of the previous studies attempted to control these variables as independent variables. Therefore, it is not certain yet which system characteristics are the most important components of operator's trust, and can affect the greatest change in trust. In prior studies (e.g., Muir & Moray, 1996; Lee & Moray, 1994), only reliability was controlled as an independent variable; other system

characteristics were subjectively assessed by participants. Recall that the reliability in these studies indicates the differences between the values the automated system actually used in controlling and the entered or reported input/output, respectively. Thus, the effect of each system characteristic of trust on the operator's trust in automated system remains unknown. Therefore, controlling the system characteristics may reveal the importance of these characteristics in contributing to different levels of operator's trust in automated systems.

Second, there has not been direct comparison made between the operator's trust in automated systems and the operator's judgment performance. While it is assumed that appropriate levels of trust lead to better performance, this has not been explicitly tested. Based on prior research in human trust in automated systems, it is hypothesized that the degree to which the human operator will rely on the decision-aid can be affected by the operator's trust in the decision-aid, which in turn is based on many characteristics of the decision-aid, including its reliability, validity, understandability, and utility. This reliance may lead, in turn, to better performance.

Third, in the domain of previous research, an automatic controller was extensively used to test the effect of reliability on the operator's trust in the automation. However, none of the research has used an automated decision-aid. The difference lies in the level of automation discussed at the beginning of this chapter. That is, in an automatic controller, the human operators have to engage in the automatic mode to know whether the automation performs the task well. However, an automated decision-aid makes the estimates of the environmental state constantly available to the human operator. The human operator has the right to either approve or veto the decision-aid's suggestion. In this way, the human operator has more chances to understand the behavior of the automated decision-aid without engaging exclusively with the automated decision-aid. This may make it more difficult to find out whether the human operator actually relies on the decision-aid or not. On the other hand, this may facilitate operator's calibration of trust.

However, despite the apparent difference between the automatic controller and the automatic decision-aid, it is hypothesized based on the previous research that the decision-aid will be trusted more when it produces estimates consistently and precisely, which are valid, and when the human operator understands the decision-aid better. One study, by Bisantz and Seong (2001), examined the effect of different types of failure of automated decision-aids on human performance. They identified different types of failures of the automated decision-aid within the adversarial aided decision-making environment, i.e., the automated decision-aid attacked by the other forces. Their results showed that different types of failure induced human operators emphasizing different aspects of the decision-aid.

2.8 Summary

In this chapter, literature reviews from human-centered automation, trust in interpersonal relationships, and trust in human factors perspective were provided first. Based on this literature review, several interesting research issues can be investigated to

increase our understanding of human operators interacting with automated systems. To discuss these issues, general model of automated decision-aids, trust, and decision-making performance was provided to describe the general conception of the environment where human operators interact with automated decision-aids to make judgments about the state of the environment. In this general description, the relationships between the three components were identified, which were to represent the level of performance of each judgment agent and to compare these judgment agents. These indices are important to identify the relationship between the operators' judgment performance and the use of the decision-aid. Based on these two relationships, operators' trust in the decision-aid can be interpreted in a normative way by linking these indices together. Discussion on this will be provided in the next chapter.

Studies on trust from the interpersonal perspective and from the human-machine interaction perspective were provided and summarized in the previous section. Also, the role of trust in human operators interacting with automated decision-aids was provided to describe the potential effect on operators' judgment and decision-making performance within the general research framework. Based on this framework, several research questions were raised to understand the effects of "imperfect" automation on operators' judgment performance and trust in conjunction with automated decision-aids. These questions are about the effects of different characteristics of automated decision- aids on operators' judgment performance and decision-making, and trust in such systems.

In addition to these research questions, another research concern in this study is to model human judgment with automated decision-aids. Having the entire framework of this study in mind, which is the effect of "imperfect" automated decision-aids on operators' judgment performance and trust in such systems; it can be improved if a model of human judgment with an automated decision-aid can represent a variety of characteristics of the automated decision-aid. The next chapter discusses this modeling framework and the development of a model of human operators' judgment with automated decision-aids.

3.0 MODELING FRAMEWORK

3.1 Modeling Literature Overview

Having discussed the importance of controlling the various characteristics of systematic or human operator's trust, this chapter presents another important research goal: developing a conceptual framework from which to study human decision-making in concert with automated decision-aids. It is based on several theoretical models and empirical studies described in the preceding chapter, including the Lens Model, and models studies of human trust. Specifically, this conceptual framework applies to the case in which human operators are making judgments or assessments of a situational state (e.g., assessing a level of threat in a military situation or making a medical diagnosis). To make this judgment, they have to assess informational cues as well as recommendations of state estimates from a decision-aid. The human operator in this case is not mandated to rely solely on the judgment of the decision-aid. It is at the human operators' discretion whether or not to rely on the situational cues or recommendations from the decision-aid. It should be iterated that this is only one of several potential styles of interaction between a human and automated system, as was noted in Chapter 2.

3.2 Lens Model Framework

The Lens Model with its extensions is a modeling framework that captures both the actual environment that people make judgment about, and the manner in which humans make judgments (Cooksey, 1996). The model is the primary conceptual and methodological tool brought forward from Brunswik's (1952) idea of probabilistic functionalism. This theory indicates the functional relationship between an organism and the environment, and states that the relationship is necessarily probabilistic or uncertain in nature. As shown in Figure 11, the focal point of the modeling framework is how the information available to the human judges (the cues, labeled $X_1, \ldots X_4$ in Fig. 11) is related to the actual environment state (Y_e), and how the information is used that to make judgments (Y_s) about the unknown environment. Several indices have been developed to evaluate the judges' performance level, namely environmental predictability, cognitive control, achievement, linear knowledge, and unmodeled knowledge. These parameters will be discussed in the next section.

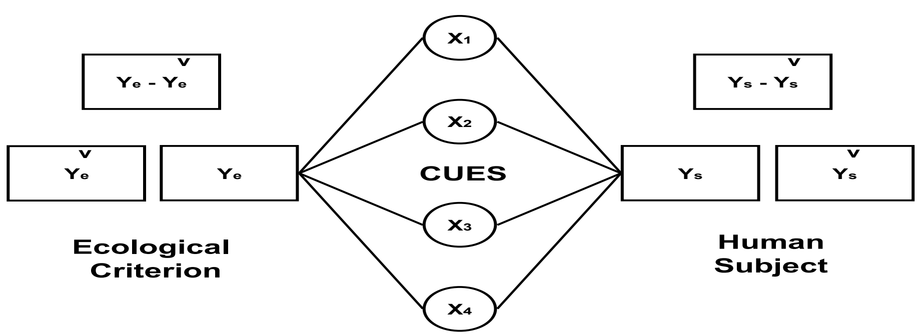

Figure 11: The Generic Lens Model and Double System Design

There are several different models within the Lens Model framework. These can be used to model different problem settings. Three of these model designs are discussed here to both explain fundamental concepts of the Lens Model, and to examine the possible application of the model to describe human operators' judgment performance along with their use of automated systems.

First, the single system design is the simplest design methodology for studying human judgment. As shown in Figure 12, a set of cue profiles (from X_1 to X_4) representing the unknown ecological criterion (Y_e) is available for the human judge to process. The ecological criterion is shown in dotted lines because for this case, its' value is unknown. For this reason, it is not possible to know how well the human judgments correspond to the ecological criterion. Once the human makes judgments based on the informational cues, the types of information that can be captured to evaluate the judgment policy are: the weighting scheme applied to the cues; the nature of the function form relating each cue to the judgments made; the degree of cognitive control which the judge exercises over the application of the policy; and finally the judge's organizing principle for integrating cue information. Hammond & Smith's (1975) study concerned planning for the future of a major U.S. university. The overall goal was to obtain views by various faculty member and academic deans. Therefore, it was not possible to know or measure what the criterion in the task ecology should or would be, thus the study was constrained to employ the single system methodology to obtain judgments of scenario desirability.

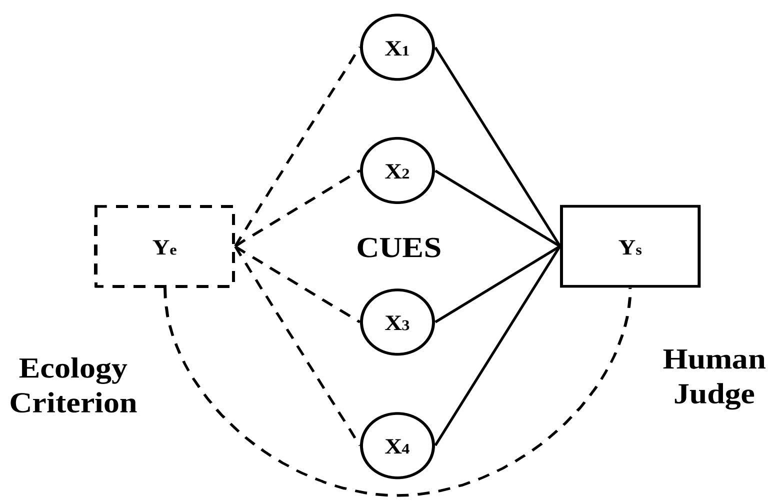

Figure 12: Single Lens Model System Design

An extension of the single system design, in which, the ecological criterion is known and available, is called double system design, shown in Figure 11. The double system design represents the typical Lens Model paradigm when a judge's cognitive system is explicitly compared to a known task or ecology system. Therefore, the double

system design overcomes the important limitation of the single system design, in that the interrelationships between the subject's cognitive system and the task system can be examined. This type of design was applied to represent how people learn to make judgments about a criterion through acquiring an understanding of the properties of a judgment task, so called multiple cue probability learning (MCPL). Bisantz (1997) used the double system design to represent a judgment task to determine the identity of an aircraft given a number of informational cues.

Next is the general n-system design. Figure 13 depicts an n-system Lens Model design where multiple judges make judgments about the environment based on an identical set of cues. The n-system design can be used to model a situation where the study focuses on a comparison of the judges. Basically, the n-system design is the logical extension of double system design to include multiple judges. Typically, the ecological criterion is unknown, or not of interest to the researcher. Hammond & Adelman (1976) used the n-system design methodology to model a dispute between the Denver City Council and the Denver Police Department regarding the type of ammunition the police should employ in their handguns.

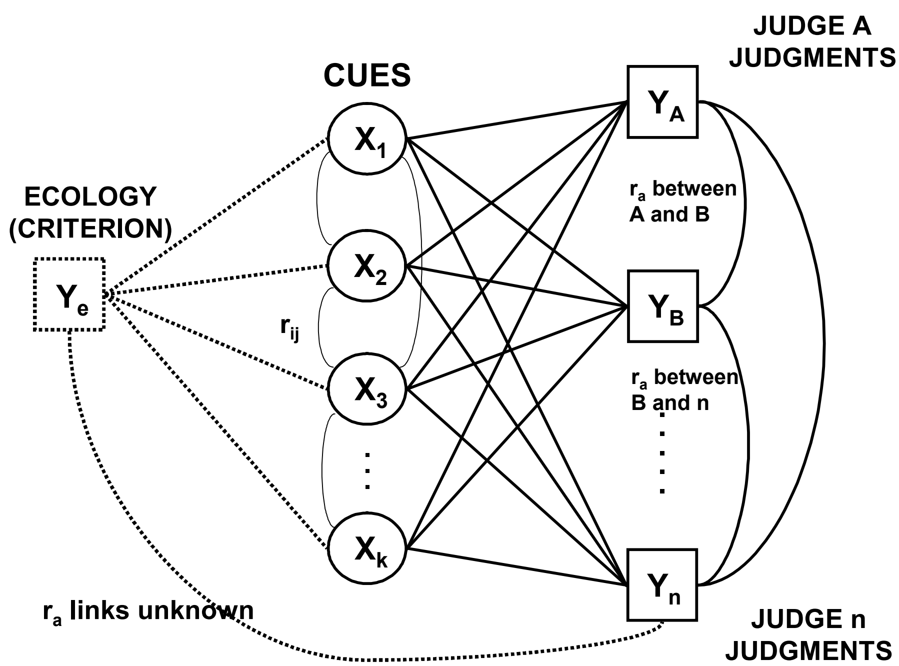

Figure 13: n-system Lens Model Design

While not frequently employed in judgment research, hierarchical judgment design is another extension of the Lens Model in which there exists multiple layers of the judgment process between the ecological criterion and the human judge. Figure 14 depicts a representation of a hierarchical system design with two layers of judgment processes and multiple sub-domains cues. First level judgments are made based on each subset of cues. These judgments, in turn, become cues for the next judgments, and so on until it reaches the final judgment. Thus, the cues in the middle between the first-order

cues and the final judgment become both judgments based on the lower set of cues as well as cues for the next higher judgments. Hammond & Smith (1975) used the hierarchical system design for investigation of planning judgments where alternative futures for a major university were evaluated. The overall goal was the mix of six planning area emphases desired by various faculty members and academic deans. The six areas were further subdivided into differing sets of objectives which could vary in their extent of contribution to each area.

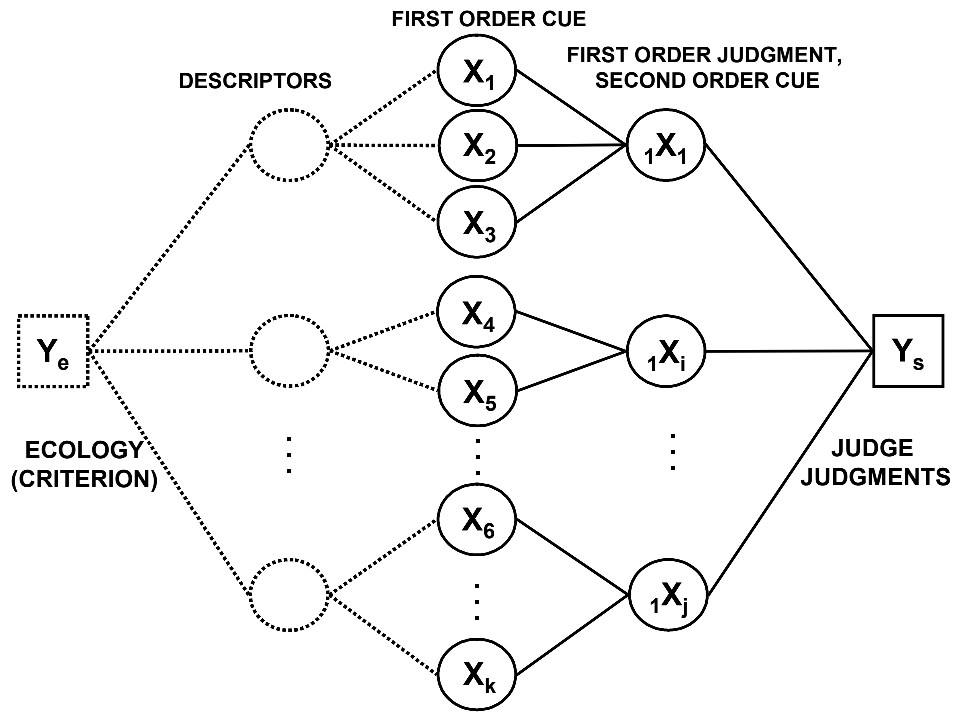

Figure 14: Hierarchical Lens Model where, k = Number of Cues in Total, j = Number of First Order Judgment which is the same as the Number of Second Order Cues

So far, four different modeling designs in the Lens Model framework have been reviewed. A discussion on applying the Lens Model to the current research issue will be provided later in this chapter. Briefly, a Hybrid Lens Model is created by integrating the characteristics of the n-system design and hierarchical Lens Model design. The Hybrid Lens Model is based on the n-system design in which the ecological environment is known and different judgment systems make judgments about the environmental states based on a set of cues. The Hybrid Lens Model is also a hierarchical Lens Model in that the outputs from the automated system are provided to the human judges to be considered in their final judgments about the environmental states.

3.2.1 Comparing Systems: The Lens Model Perspective

Cooksey (1996) provides an extensive discussion on the Lens Model framework and their associated parameters, which can be used to measure and capture people's judgment and decision-making policies. The focus in this section is on these statistical

indices which arise out of this Lens Model conceptualization. The initial discussion will consider the Lens Model as implemented in the double-system design, as shown in Figure 15. This will lead to the discussion on the statistical indices of the Hybrid system design that was used in this study.

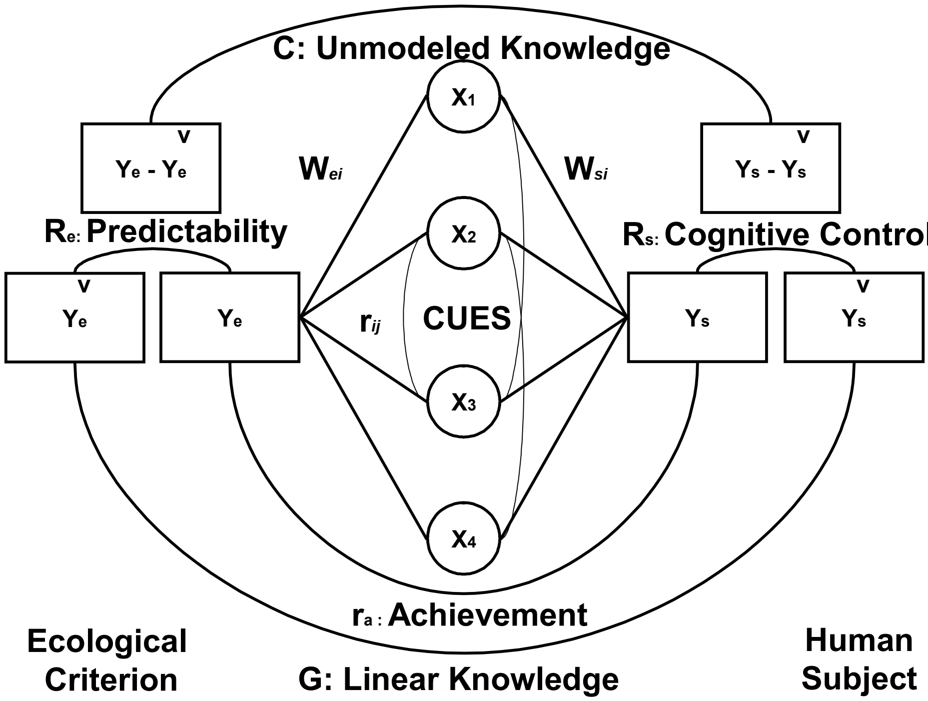

Figure 15: The Lens Model with the Statistical Parameters Superimposed

3.2.1.1 The Standard Lens Model Equation

Lens Model parameters can be computed based on the collection of data that includes actual judgments, environmental criteria, and cue values. The following parameters indicated in Figure 15 can be computed:

Achievement: Achievement (r_a) indicates the level of performance to the extent which the participant's judgments are in accordance with the environmental criteria. This is calculated by correlating the two fundamental components, the actual environmental criteria and the actual judgments. This correlation represents the extent to which judgments and criterion values agree. A high level of achievement represents the situation where environmental values and judgments are close together.

Cognitive Control: Once the judgments are made, a multiple regression analysis between the set of cue information provided and the judgments made can be used to generate estimates for the judgments. The correlation between the actual judgments and the estimates of those judgments is used as an index to represent the degree to which the participants are able to implement their knowledge or judgment policies into the judgment task. While this index of cognitive control (R_s) is somewhat similar to the

degree of consistency in executing the policy, Hammond et al. (1975) claimed that there existed a technical distinction between cognitive control and consistency. When the same judgment profiles whose set of cue information is identical are repeated within a set of trials, the index of consistency indicates the extent to which operators can make the same judgments over those repeated profiles. This can be very critical in a situation where any devices providing information to the operators are consistently incorrect. In a military domain for example, suppose that the informational cues measured by a group of sensors consistently indicate that the unknown objects are hostile whose true identities are otherwise. Cooksey (1996) summarized a procedure to statistically separate the level of consistency from cognitive control, which will be discussed in detail in the next chapter.

Environmental Predictability: When the true states of the environmental criterion are known or of interest to the research, one can examine the relationship between the set of cue information provided and the environmental criterion, termed environmental predictability (R_e). The concept and the method to calculate the environmental predictability are same as calculating cognitive control of a subject. Environment predictability measures how well the set of cue information corresponds to the environmental states. Basically, this index represents how predictable the set of information is in terms of the environmental states.

Linear Knowledge: Often called *G*, linear knowledge measures how well the predictions of the model of the human judge correspond to predictions of the model of the environment. That is, this index shows how well the linear component of the human's judgment corresponds to the linear component of the environment. A high level of linear knowledge indicates that both models are similar and therefore predictions will be more likely to coincide. It suggests that the human judge captures the linear structure of the environment very well. The Linear Knowledge, *G*, is mathematically defined as the correlation between predicted criterion values and predicted judgments.

$$G = Corr(\hat{Y}_e, \hat{Y}_s) \qquad \text{Equation 1}$$

Unmodeled Knowledge: Often called *C* to indicate configurality, unmodeled knowledge captures correspondence between systematic variations in both models which are embedded within the residual terms for each model. Therefore, this index captures the degree of non-linearity to the extent that both models share the same non-linear components. However, interpretation of unmodeled knowledge is somewhat different from the previous interpretation of linear knowledge. A high value for unmodeled knowledge indicates the human judge applies the knowledge of prediction on unmodeled relationships in the environment with reasonable accuracy. Note that all of these Lens Model parameters are the parameters from the regression and correlation analyses. Unmodeled Knowledge, *C*, is defined as the correlation between the residuals of the environment and of the human judgment model.

$$C = Corr((Y_e - \hat{Y}_e), (Y_s - \hat{Y}_s)) \qquad \text{Equation 2}$$

Hursch, Hammond, & Hursch (1964) derived the first Lens Model Equation (LME) which provided the integrative statistical formulation needed to accomplish the person-ecology comparison task. Later, Tucker (1964) modified this equation with a simpler formulation that has become the standard form of the LME used today. The LME is

$$r_a = GR_eR_s + C\sqrt{(1 - R_e^2)}\sqrt{(1 - R_s^2)} \qquad \text{Equation 3}$$

where
r_a: achievement, R_e: environmental predictability, R_s: cognitive control, G: linear knowledge and C: unmodeled knowledge.

In Equation 3, a judge's achievement is a function of two multiplicative components. The first component of the equation is termed the linear component and represents the component of the achievement correlation which can be attributed to the explicit linear modeling. The second component of Equation 3, originally called a configural component, and later termed the unmodeled component, reflects the presence of the unmodeled knowledge (C).

The equation incorporates an interesting point where the achievement can be calculated solely based on the linear knowledge, task predictability, and cognitive control, if the unmodeled component can be assumed to be null.

In addition to the basic Lens Model parameters, an index relevant to this study is the degree to which a judge makes judgments in a consistent manner. Consistency defined as the degree to which a judge makes consistent judgments on the profiles that have the same set of informational cues. It sounds similar to the index of cognitive control. Hammond et al. (1975) claimed that a technical distinction between cognitive control and consistency with respect to the interpretation of R_s should be made. This distinction between two parameters will be discussed in the next chapter in more detail because it provides important control over the independent variables in this study.

Cooksey (1996) explained the procedure for statistically separating cognitive control from consistency by using the "lack-of-fits" test by Draper and Smith (1966). First, it is necessary to test whether consistency must be separated from cognitive control. Draper and Smith (1966) provided a procedure, called "lack-of-fits," which could identify whether the policy model accounts for all the systematic variation in the individual's judgment. The test for "lack-of-fits" is basically whether the total variance of the individual's judgment (S_Y^2) is greater than the total variance of individual's judgments for the repeated profiles (S_e^2). If "lack-of-fits" is not demonstrated (meaning that the total variance of the individual judgments for the repeated profiles is equal to or greater than the total variance of the individual's judgments), then the assumption of Y_j

and $\hat{Y}_j$ coinciding is tenable and one can use R_s to index both cognitive control and consistency. Otherwise then, R_s only measures cognitive control and one needs a separate index for consistency.

3.3 Lens Model of Human Operators with Automated Decision-aids

Recall that in the previous section, the type of automated decision-aid of concern in the present research was indicated to be one in which the human operator, as well as the automated decision-aid, both have access to informational cues and produce judgments. The human operator is free to incorporate the output of the automated decision-aid into their judgments. Based on this description of the automated decision-aid and its relationship with the environment and the human operators, the Lens Model can be applied (shown in Fig. 9). Figure 16 shows a Lens Model which includes the judgments produced by the automated decision-aid (ADA), the human operator's judgments in conjunction with the automated decision-aid, denoted HO_{ADA}, and the human operator without the automated decision-aid, denoted *HO*. The task of the operator is to provide estimates about the state of the environment based on the information provided, denoted as X_i. In this figure, there is a dotted arrow link (between *ADA* and HO_{ADA}) representing the input from the automated decision-aid to the human operator. This link represents not only the informational output or state estimate provided by the decision-aid, but characteristics of the decision-aid itself. That is, the human operator not only considers the direct estimates from the automated decision-aid, but also considers the characteristics of the automated decision-aid such as its reliability and validity. Therefore, the automated decision-aid, as a whole, is considered by the human operator in determining whether or not to take the estimates into consideration in the final judgment. Also, the link is shown as a dotted line to indicate that it is the operator's decision to ask to see the estimates from the automated decision-aid. The locus of authority over the decision-aid's estimates is the human operator. That is, when the operator's trust in the decision-aid decreases, the operator may not seek the estimates from the decision-aid.

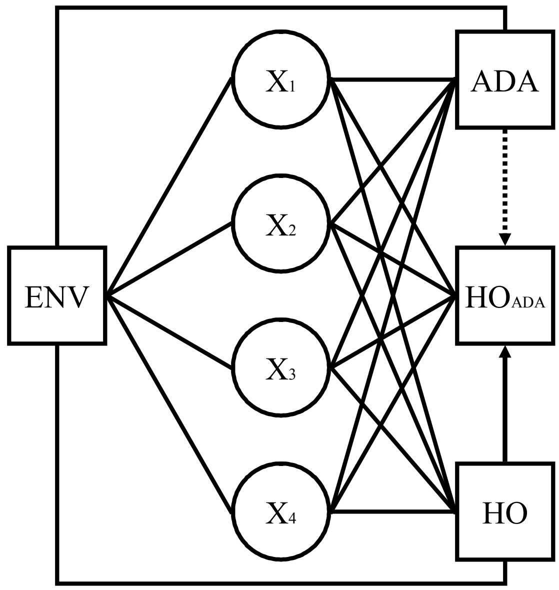

Figure 16: Application of the Lens Model - ENV: Environment, HO: Human Operator, ADA: Automated Decision-aid

Figure 17 represents the Lens Model using three connected models, in order to better illustrate the comparisons between the models, annotated with several important relationships in comparing the judgment performance. While the set of information cues, which are provided to the automated decision-aid and the human operator, are the same, as shown in the top and middle Lens Model in Figure 17, the estimates from the decision-aid become an information cue in the Lens Model of the human operator acting in conjunction with the automated decision-aid. This adds another cue in the bottom Lens Model in Figure 17. It is important to emphasize that while the model assumes that the same set of information is available to both the human operator and the automated decision-aids, the model itself makes no claims regarding the degree or manner in which those informational cues are utilized by either of the two judgment systems. In the most extreme case, it is entirely possible within this modeling framework for the automated decision-aid to rely on a completely different subset of the available cues than the human operator. More likely, each relies on some overlapping set of cues, to different degrees. The evaluation of the manner in which cues are utilized by either is, in fact an empirical question, measured through the Lens Model parameters.

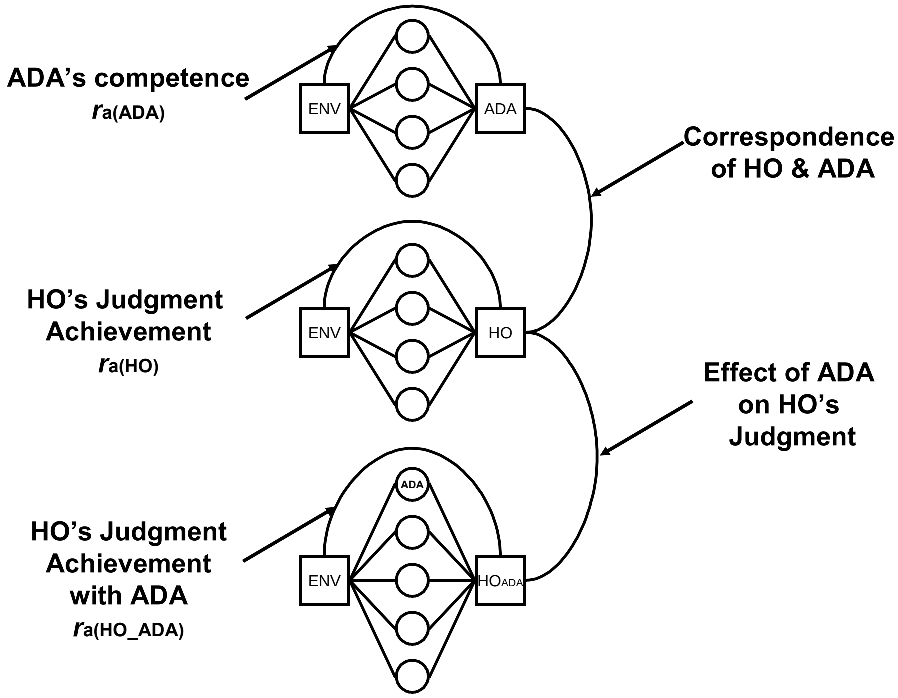

Figure 17: Connected Three Lens Models of an Automated Decision-aid, Human Operators' Judgment with or without the Automated Decision-aid showing the Correspondence and the Effect of ADA on the Judgments

The arcs in each model connecting both sides of the system components refer to general achievement indicating the level of performance of each judgment agent. First, the arc in the top of Figure 17 represents the degree to which the automated decision-aid can estimate the environmental state. That is, the arc represents the degree to which the automated decision-aid's estimates are correlated with the environmental state. These arcs are computed by correlating the true environmental state and the judgment of the right side of each Lens Model, i.e., the estimates from the automated decision-aid in this example.

Second, comparing the decision-aid's estimates with the human operator's judgments can provide important information to the extent which both judgment agents are similar to each other in estimating the environment. It is labeled here as the correspondence between the human operators and the automated decision-aid.

Third, when the estimates from the automated decision-aid are available and provided to the human operators, the level of human operators' performance may be affected. This is indicated by the arc between the human operator and the human operator with the automated decision-aid, denoted *HO* and ***HO_ADA,*** respectively, in Figure 17.

These Lens Models may be able to illuminate important characteristics of human operators' trust in the decision-aid and their resulting strategies in considering the decision-aid's estimates. For example, assume that the automated decision-aid produces

accurate estimates of the environmental states. This should result in a high level of performance by the human operator with the decision-aid, provided that the human operator trusts the decision-aid's estimate and decides to rely on it. In addition, if the human operator is able to make appropriate judgments, the level of correspondence between the human operator and the decision-aid will be high and also all the levels of performance will be high.

In this section, an overall conceptual framework to study human operators' trust in automated decision-aids and judgment and decision-making performance within the context of the Lens Model has been discussed. Important relationships regarding the human operator's trust in the automated decision-aid, and the indices from the Lens Model established have been briefly discussed. The following section examines components of each of the models in detail. Further details on each Lens Model will be discussed in the following section.

3.4 Details on the Lens Models

A general discussion of the conceptual model of human operators' judgment and trust in automated decision-aid has been provided in the previous section. The following section describes the three related Lens Model and their associated parameters in detail.

3.4.1 Lens Model of Human Operator's Judgment

The model of a human operator's judgment of the state of the environment is a prototypical example of the application of the Lens Model framework. In this model, shown in Figure 18, the human judgment agent makes a judgment (Y_{HO}) based on the informational cues provided from the environmental state (Y_{env}). Predictions of the environmental state and the operator's judgment ($\hat{Y}_{\mathrm{env}}$ and $\hat{Y}_{\mathrm{HO},}$ respectively) can also be found based on their relationships with the informational cues. The model captures the human operator's judgment performance ($r_{\mathrm{a(HO)}}$), the ability to implement the mental model of judgment (cognitive control, denoted as R_{s}), and the components of judgment performance which can be captured in a linear model (G_{HO}) and components of judgment which can not be captured in a linear model (C_{HO}). Additionally, R_{e}, environmental predictability, represents the linear predictability of the environment. $W_{\mathrm{i(HO)}}$, the operator's weighting scheme on the informational cues, can be used to characterize the operator's apparent strategy in making the judgment.

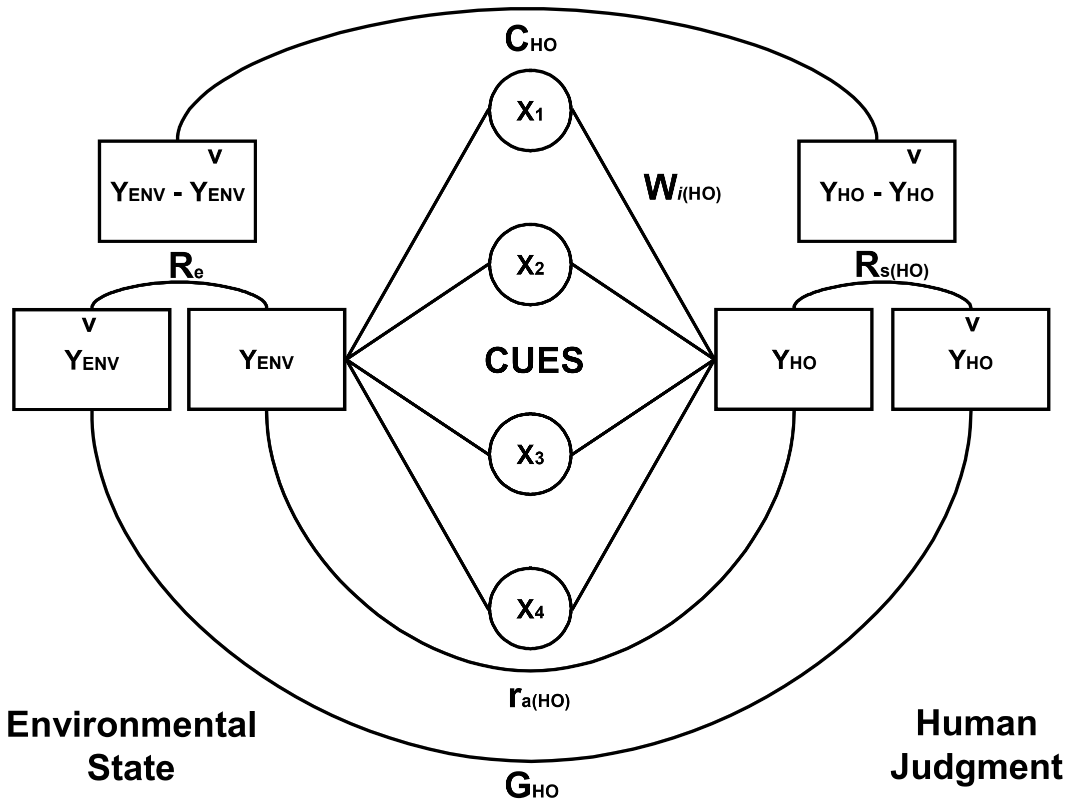

Figure 18: Lens Model of Human Operators' Judgment without an Automated Decision-aid

Table 6 explains the Lens Model parameters within the context of Figure 18. Since the Lens Model of human operators' judgment of the environmental state is a typical example, the description is very straightforward in accordance with the definitions of those Lens Model parameters. For example, the achievement parameter, denoted $r_{a(HO)}$, indicates the degree to which the environmental states are in accordance with the human operators' judgment strategy calculated by the correlation model. Since descriptions of the Lens Model parameters are provided in the previous chapter, they will not be discussed in detail in this section.

Table 6. Description of the Lens Model Parameters of Human Operators' Judgment without an Automated Decision-aid. LMP: Lens Model Parameters

LMP	Denoted	Description
$\boldsymbol{W}_\mathrm{i}$	$\boldsymbol{W}_{\mathrm{i(HO)}}$	HO's weighting scheme
$\boldsymbol{r}_\mathrm{a}$	$\boldsymbol{r}_{\mathrm{a(HO)}}$	HO's achievement; judgment performance
$\boldsymbol{G}$	$\boldsymbol{G}_{\mathrm{HO}}$	HO's knowledge about the environment
$\boldsymbol{C}$	$\boldsymbol{C}_{\mathrm{HO}}$	Unmodeled knowledge
$\boldsymbol{R}_\mathrm{e}$	$\boldsymbol{R}_\mathrm{e}$	Environmental predictability
$\boldsymbol{R}_\mathrm{s}$	$\boldsymbol{R}_{\mathrm{s(HO)}}$	Cognitive control

3.4.2 Lens Model of Automated Decision-aid Performance

Applying the Lens Model framework to the automated decision-aid is similarly straightforward (see Fig. 19). Here, the state of the environment is presented via the cues (X_i) shown in the middle, which the automated decision-aid uses to produce the estimates. This set of informational cues is the same as those in the Lens Model of the human operator's judgments. The automated decision-aid calculates the estimates of the environmental state that will be available for the human operators to consider in order to render the judgment.

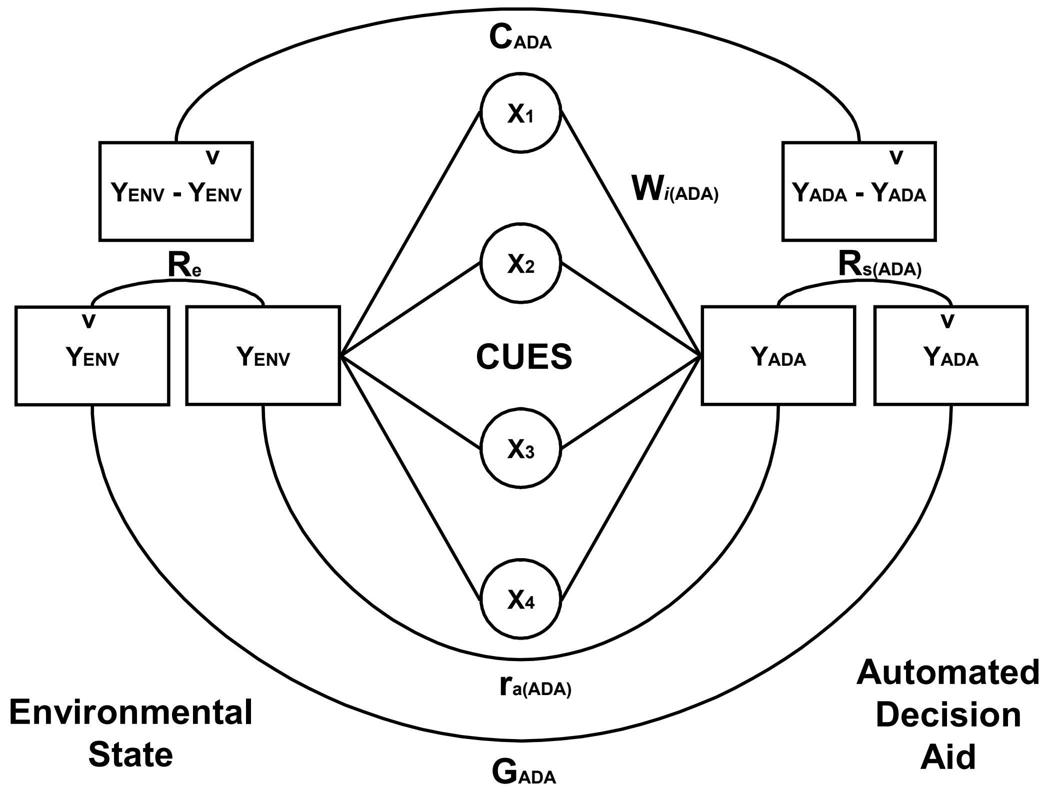

Figure 19: Lens Model of an Automated Decision-aid

When the Lens Model is applied to model the automated decision-aid's performance estimating the environmental state, the Lens Model parameters take somewhat different meanings. In this case, the Lens Model parameters can be used to evaluate various aspects of the automated decision-aid. Table 7 shows the Lens Model parameters with its description of each parameter when it is applied to the automated decision-aid. For example, the parameter to represent the human operators' performance, achievement denoted as $\boldsymbol{r}_{a}$, becomes an index to evaluate the decision-aid's performance. That is, the index represents the degree of ability or competence to which the automated decision-aid can estimate the state of the environment. Also, it is interesting to note that another parameter, linear knowledge (G_{ADA}), can be used to evaluate the automated decision-aid performance and competence. Generally, linear knowledge in the Lens Model describes the degree to which the predicted model of the environment and the predicted human operator's knowledge implementation are in accordance with each other. However, when it is used to describe an automated decision-aid, this can represent the knowledge that is implemented in the automated decision-aid. Because this parameter represents the degree of correlation between the predicted environments and the predicted automated decision-aid's knowledge base, this is essentially the same as the knowledge base implemented in the automated decision-aid. Thus, this parameter can be an index which describes the degree to which the automated decision-aid can perform the task for situations or represent the states of the environment that have not been expected in the design stage. This aspect can represent the degree of the decision-aid's sensitivity to a real and unforeseen environment. Similarly, unmodeled knowledge, denoted as C_{ADA},

represents the degree of knowledge not implemented in the automated decision-aid. Therefore, both parameters, linear and unmodeled knowledge, can be utilized to evaluate the degree to which the automated decision-aid's knowledge base fits the actual environment.

Table 7. Lens Model Parameters for the Automated Decision-aid. * see text for detail

LMP	Denoted	Description
$\boldsymbol{W}_{\mathbf{i}}$	$\boldsymbol{W}_{\mathrm{i(ADA)}}$	ADA's designed scheme
$\boldsymbol{r}_{\mathbf{a}}$	$\boldsymbol{r}_{\mathrm{a(ADA)}}$	ADA's achievement; ADA's competence
$\boldsymbol{G}$	$\boldsymbol{G}_{\mathrm{ADA}}$	Knowledge that is implemented in the ADA
$\boldsymbol{C}$	$\boldsymbol{C}_{\mathrm{ADA}}$	Knowledge that is not considered in the ADA
$\hat{\boldsymbol{Y}}_{\mathbf{e}}$	$\hat{\boldsymbol{Y}}_{\mathrm{env}}$	Predicted environment
$\boldsymbol{R}_{\mathbf{e}}$	$\boldsymbol{R}_{\mathrm{e}}$	Environmental predictability
$\boldsymbol{R}_{\mathbf{s}}$	$\boldsymbol{R}_{\mathrm{s(ADA)}}$	Cognitive control (Consistency*)

The parameter typically known as cognitive control, denoted as $R_{\mathrm{s(ADA)}}$, is also interesting for two related reasons. First, this parameter indicates the degree to which the automated decision-aid performs consistently in the actual environment. As indicated, cognitive control is the correlation between the actual judgment and the predicted judgment. From an automated decision-aid perspective, this can be seen as how consistent the automated decision-aid is in making judgments based on the information provided for the estimates. The other reason why this is interesting is that because by measuring the consistency of the automated decision-aid in producing the estimates, it measures an important factor that has been hypothesized to affect the level of human operators' trust in automated decision-aids: the reliability of performance of the automated system. Automated decision-aids that behave consistently should be trusted more than inconsistent ones.

3.4.3 Lens Model of Human Operator's Performance with Automated Decision-aid

When an automated decision-aid is introduced to support human operators' judgment and decision-making, one way to consider the automated decision-aid is as another informational cue for the human operators to consider rendering the final judgment. Because the automated decision-aid integrates the information measured from the state of the environment, it may be weighted more than other information cues. However, the extent to which the human operator relies on the automated decision-aid, or any other available information cue, is an open question. For instance, as indicated above, trust in the automated decision-aid may impact an operator's decision to rely on it.

Figure 20 shows the Lens Model of human operators with an automated decision-aid. In this figure, the automated decision-aid is added between the environment and the human operator, which means that it is considered as an informational cue similar to other information cues. In a judgment and decision-making framework, an automated decision-aid can be thought of as another cue that is available to the human operator to consider, but with a weight that reflects the human operator's reliance in the automated decision-aid. As will be discussed later, reliance on the automated decision-aid may increase as the operators trust increases. However, it is not necessarily true that the automated decision-aid will be emphasized more than other informational cues.

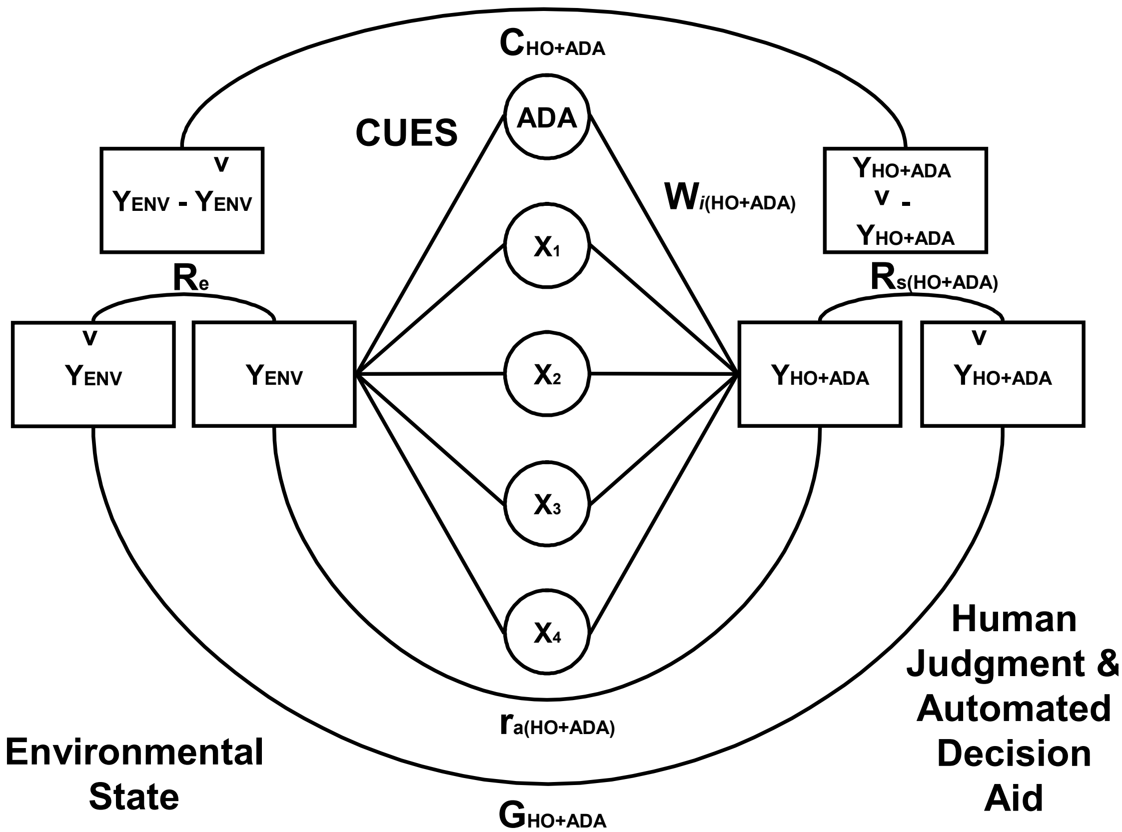

Figure 20: Lens Model of Human Operator's Judgment with an Automated Decision-aid

Table 8 shows description of the Lens Model parameters of the Lens Model of human operators with an automated decision-aid.

Table 8. Lens Model Parameters for Human Operators' Judgment with an Automated Decision-aid

LMP	Denoted	Description
$\boldsymbol{W}_{\text{i}}$	$\boldsymbol{W}_{\text{i(HO\_ADA)}}$	Weighting scheme by HO with ADA
$\boldsymbol{r}_{\text{a}}$	$\boldsymbol{r}_{\text{a(HO\_ADA)}}$	Composite achievement
$\boldsymbol{G}$	$\boldsymbol{G}_{\text{HO\_ADA}}$	Composite linear knowledge
$\boldsymbol{C}$	$\boldsymbol{C}_{\text{HO\_ADA}}$	Composite Unmodeled Knowledge
$\boldsymbol{R}_{\text{e}}$	$\boldsymbol{R}_{\text{e}}$	Environmental predictability
$\boldsymbol{R}_{\text{s}}$	$\boldsymbol{R}_{\text{s(HO\_ADA)}}$	Composite cognitive control

3.5 Discussions on Models

Separate discussions on each Lens Model have been provided. Comparisons can be made to describe the relationships between the components which are shown in Figure 9. The relationships between the automated decision-aid or the human operator and the environmental state have been discussed through the Lens Model framework. Recombining the three separate models makes the relationships between them explicit as well as illustrating the nature of the model as a combination of n-system and hierarchical Lens Model.

Figure 21 depicts an overall Hybrid Lens Model of human operator's judgment with the automated decision-aid. In an n-system Lens Model design, multiple judgment agents exist, and their judgment performance can be compared against each other. The model shown here is partly an n-system design in that the information cues are provided to the two judgment and decision-making agents: the decision-aid and the human operator. Like one of the purposes of the n-system Lens Model design, comparing the judgments between the two judgments and decision-making agents can be accomplished.

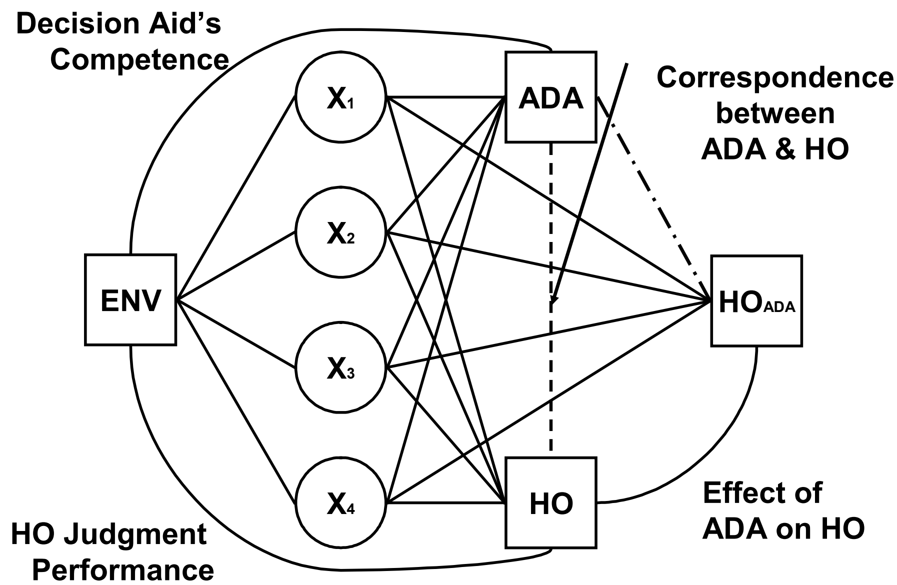

Figure 21: Hybrid Lens Model of Human Operators' Judgment with Automated Decision-aids

In a hierarchical system design on the other hand, judgments of the first level agent become cues to the second level judgment agent. This model has an aspect of a hierarchical system design in that the human operator and the automated decision-aid can be seen as making independent judgments, which are then combined to produce an aided judgment. Therefore, this is a hybrid model, containing aspects of both n-system and hierarchical model.

Figure 21 also includes several measures that represent various relationships within the domain of interest. First, is the human operators' judgment performance, represented by the arc between the environmental criterion and the human operators, and denoted "HO Judgment Performance," which can be measured by achievement in the Lens Model term. Additionally, the decision-aid's competence, represented by the arc between the environmental criterion and the decision-aid, can also be measured by correlating the two. Finally, the effect of the automated decision-aid on human operators' judgment performance, represented by the arc between the ADA and HO, can be measured by also correlating the automated decision-aid's estimates and the human operators' judgments. Methodologically, this can be achieved by employing a within-subject experimental design.

3.6 Application of the Modeling Framework to the Study of Trust

3.6.1 Measuring & Manipulating Reliability, Validity, & Understandability

Within the modeling framework presented here, several factors previously hypothesized to affect trust can be systematically measured, and experimentally manipulated. The reliability of an automated decision-aid can be measured by the

regression model indicating the consistency with which the decision-aid makes judgments (e.g., $R_{s(ADA)}$). Second, the validity of the automated decision-aid can be measured by $\boldsymbol{r}_{a(ADA)}$, labeled in Figure 9 as the Degree of Aid Competence.

Third, the understandability of an automated system, or the degree to which human operators understand why and how the system is behaving, has been identified as a contributing factor to human operator trust. Within this modeling framework, providing an operator with information regarding reliability and validity (that is $R_{s(ADA)}$ and $\boldsymbol{r}_{a(ADA)}$, respectively) may positively contribute to understandability. Note that these types of information concern the relationship between the environment, the informational cues, and the automated decision-aid's estimates. This is important because such information is identical to some forms of cognitive feedback that have been shown in previous studies to increase the level of operator's judgment performance.

Some studies have investigated the role of feedback on judgment performance. For instance, Balzer, Doherty, & O'Connor (1989) performed an extensive literature review on different types of feedback that may have potential effects on operator's judgment and decision-making performance. These include task information, cognitive information, and functional validity information feedback. First, task information includes two relationships; one relationship between the environmental states and the set of cue information, and the other relationship between the environmental states and the estimated environmental states based on the first relationship described above. Basically, ecological validity information concerns statistical relationship between the environment and set of cue information provided to or made available to human operators. Thus, the Lens Model parameters about these relationships are included in this category (Fig. 22); environmental predictability (R_e), cue intercorrelation (r_{ij}), and ecological validities (W_{ei}). The second type of feedback is cognitive information. It is concerned with the other end of the entire system, from the set of cue information to the human operators at the other end. This type of feedback includes human subject's cue weighting strategy (W_{si}), and cognitive control (R_s). Finally, functional validity feedback includes the types of information about the relationships between the environment and the human subjects. Among the Lens Model parameters, this type of feedback includes achievement (r_a), linear knowledge (G), and unmodeled knowledge (C).

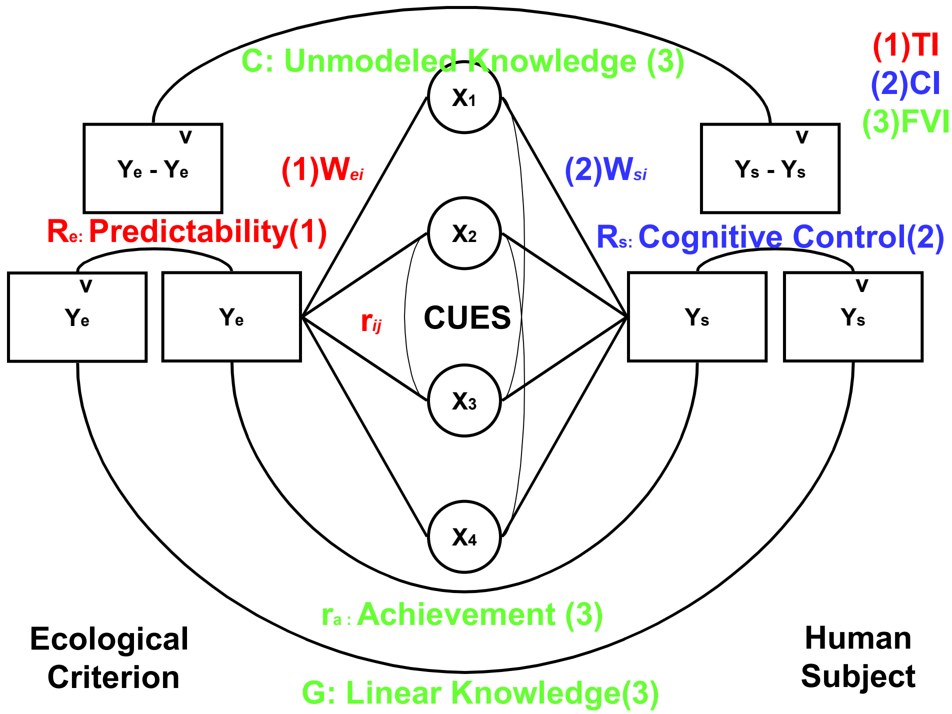

Figure 22: Lens Model with its Parameters Categorized by the Types of Cognitive Feedback

Balzer et al. (1992) performed an experiment to see the effects of different types of feedback on judges' performance in a baseball judgment task, and concluded that the task information was the most important feature contributing the level of judges' performance. It is interesting that any further feedback provided together with the task information did not contribute to the level of performance. This has practical implications in that it can allow reducing the amount of training that judges need to perform by focusing on learning the task related information rather than concentrating on discovering their judgment policies.

On the other hand, a recent study by Gattie & Bisantz (2002) investigated the effects of different types of feedback on judgment performance in the dental domain. Two groups of participants with differing with levels of expertise in the domain were provided one of those three types of feedback. The degree of linear predictability of the environment was also controlled. He found that the cognitive information was the most important information to increase for novices in terms of the Lens Model parameters. However, for experts it seemed to make no difference in the use of different types of feedback.

Thus, replacing the human judge in the Lens Model with one of the judgment agents in this study, the automated decision-aid as shown in the previous chapter, the Lens Model parameters transform into indices that can be used to interpret various aspects of the automated decision-aid. For example, cognitive control, which indicates the extent a human judge is able to execute his or her knowledge into the actual judgment

and decision-making task, can be interpreted as how well the automated decision-aid produces the estimates consistently. Basically, those Lens Model parameters can be seen as indices that can show meta-relationships between the components. Therefore, when the same notion of cognitive feedback is applied to the automated decision-aid in this study, it carries a very important link to one of the systematic characteristics of trust defined previously; understandability. That is, providing cognitive feedback of the automated decision-aid can be a very efficient approach to increase the operator's level of understanding of the automated decision-aid, which may lead to the development of trust.

Additionally, parameters in the model can used to measure aspects of performance and trust. For example, the level of correspondence between the judgment of the aided human operator (with the judgment provided by the decision-aid), labeled in Figure 9 as the Degree of Agreement, along with the degree of similarity between the unaided and aided operator judgments, may be one possible indication of the level of trust an operator has in the automated decision-aid.

3.6.2 Summary of the Three Factors

Discussions on the three factors have been provided to investigate the effects of different characteristics of the automated decision-aids on operators' judgment performance. The three variables were graphically shown in Figure 23 within the Lens Model framework. Validity indicated the relationship between the environmental state and the automated decision-aid's estimates. This index was depicted as the connecting line between the environmental state and the automated decision-aid's estimate. Reliability referred to the degree of consistency of the automated decision-aid's estimates over the profiles. This index was depicted as the arrow over the automated decision-aid's estimates. Finally, several types of information were identified to increase the operators' understanding of the automated decision-aids, including the two factors discussed above, and cognitive information. The cognitive information includes the weighting scheme over the informational cues, and the cognitive control of the automated decision-aids that coincided with the reliability factor in this study.

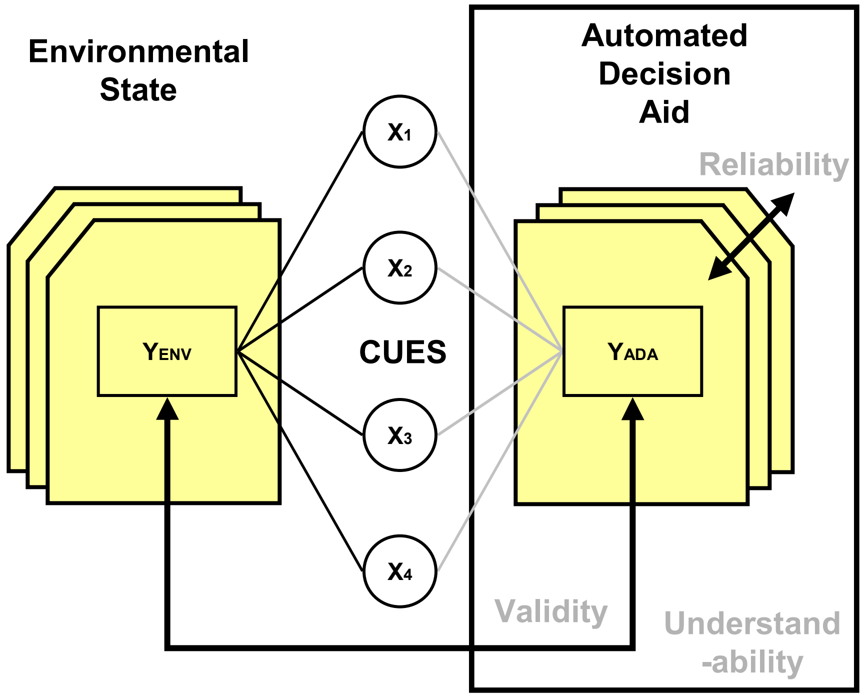

Figure 23: Representation of the Three Independent Variables with Respect to the Lens Model

Having discussed the method of controlling these variables that were hypothesized to affect operators' judgment performance, it is necessary to examine the relationships between the three components in this study, and more importantly the relationship with human trust in automated decision-aids.

3.6.3 The Relationship between Human & Decision-aid Competence & Human Trust

Based on the degree of aid competence and the degree of agreement suggested here, several expectations and hypotheses can be made.

Figure 24 depicts how the three links shown in Figure 9 are related to each other by plotting the two degrees of similarities between the components. The degree of aid competence indicates the similarity between the environmental state and the estimates from the automated decision-aid, and the degree of agreement refers to the similarity between the estimates from the decision-aid and the human operator's judgment. Since the environmental state is the ultimate target value that both the human operator and the decision-aid attempt to identify, the degree of aid competence is an absolute value. This means that the larger the term the better. On the other hand, unlike the degree of aid competence, the degree of agreement is a relative term indicating that its value should be interpreted depending on the degree of aid competence. For example, consider when the degree of aid competence is high (Quadrants III and IV). This means that the estimates from the automated decision-aid are accurate, which also means that the human's judgment should normatively match that of the automated decision-aid: a high degree of agreement (Quadrant III). In the opposite case, where the degree of competence is small,

(Quadrants I and II), the degree of agreement should also be small (Quadrant I): the human, in order to make a sound judgment, should reject the advice of the aid. Therefore, while the judgment grid can represent the normative relationships between the components, by plotting the two degrees of differences on the plane can also represent the human operator's judgment behavior in accordance with the estimates from the decision-aid in a descriptive way.

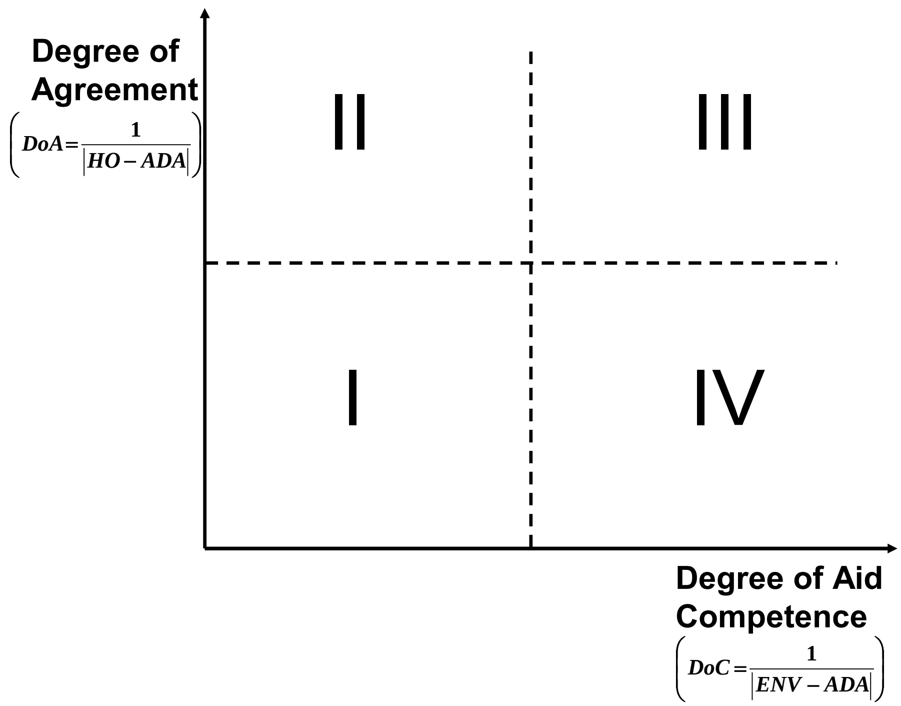

Figure 24: A Template Grid for Analyzing Operators' Judgment Performance Relative to the Automated Decision-aid, Environmental State, and Normative Level of Trust

Quadrants II and IV, are the cases where the operators fail to calibrate their levels of trust in the decision-aid appropriately. Quadrant II is indicative of the operator blindly trusting the decision-aid without knowing its real competence, and/or the operator's poor ability to generate the estimate of the environmental state. This is the case where human operators "misuse" the automated decision-aids. Similarly, the operator's trust in the decision-aid in Quadrant IV is low because of the operator's false belief in his ability to estimate the environment. This is the case where human operators "disuse" the automated decision-aids. Also, this quadrant has a special meaning when the entire sequence of the operator's judgment is plotted in this plane. Assume that the decision-aid provided inaccurate estimates of the environmental state, and has performed the task well for the last few attempts. When the operator's current judgment falls into Quadrant IV, this situation can mean that the operator's trust has not completely recovered from the previous low level of trust. This "inertia" effect has been shown in the prior research (Lee & Moray, 1992; Lerch & Prietula, 1989); that the operator's trust in automation does not

recover to the level where it should have been. Therefore, the simple two-dimensional space can be used to estimate operators' normative level of trust based on the two indices which represent some aspects of performance of both judgment agents in a discrete way (Tbl. 9). Also, it can be used to describe operators' trust when they fail to perform the judgment tasks in the normative way.

Table 9. Operator's Normative Level of Trust in the Automated Decision-aid According to the Quadrants in the Judgment and Trust Grid. O: Correct, X: Incorrect Judgment

Quadrant	DoC	DoA	ADA	HO	Performance	Trust
I	Low	Low	X	O	High	Low
II	Low	High	X	X	Low	High
III	High	High	O	O	High	High
IV	High	Low	O	X	Low	Low

Recall that causes miscalibration can also be described in terms of this modeling framework. The *C* coefficient represents the extent to which the two systems' policies in question employ similar and substantive, yet unmodeled, cues in the formation of their judgments. Also, the *G* coefficient represents the similarity of the linear models of the automated decision-aids the human operators exerted in the judgment and decision-making tasks. If it is assumed that the coefficient C is negligible, several cases can be possible in which the agreement between two judgment agents may be low: First, agreement can be low because either or both agents show low levels of cognitive control, which means that the two agents lack of ability to execute their judgment policy. However, low levels of cognitive control may indicate that it is possible for both judgment agents to agree in some of their judgments despite holding quite dissimilar policies. Alternatively, agreement can be also low despite of the high levels of two agents' cognitive control because the linear models of both agents are not similar.

Cooksey (1996) discussed the first case as "false agreement," which is shown in the top of Figure 25. With greater dissimilarity between the policies of the two judgment agents, the distance between the two means of judgments exerted by two agents is greater than when the policies are similar. The shaded region, which is the overlapping region of two judgment distributions, describes the situation where the two agents' judgments are in accordance with each other despite the largely different judgment policies. On the other hand when the judgment policies are similar, which means a higher *G*, regions of "false disagreement" (Cooksey, 1996) can be exhibited when an agent makes a judgment that would not be executed by the other judgment agent. This is shown in the bottom of Figure 25. The areas are those that cannot be covered by the other judgment distribution.

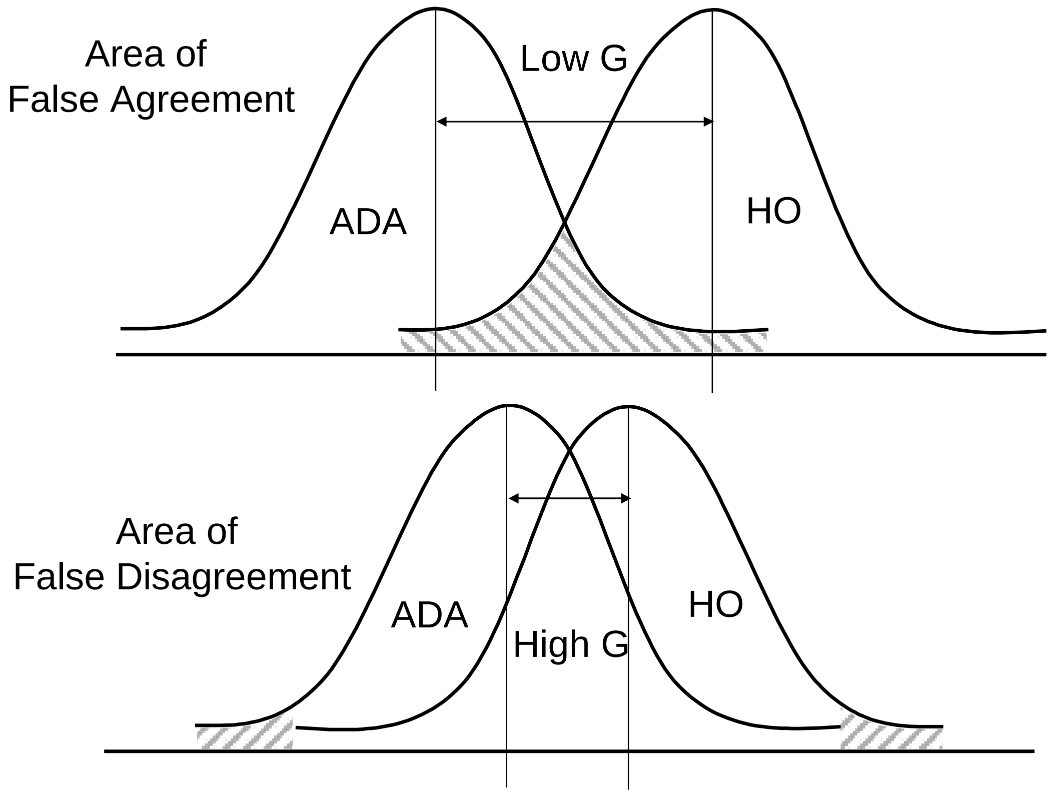

Figure 25: Regions of False Agreement and False Disagreement when Comparing Two Judgment Agents (adapted from Cooksey, 1996)

Hammond & Brehmer (1973) demonstrate that the results of false agreement or false disagreement can be quite serious, and lead to distrust and an increased level of conflict between the judgment agents. That is, these two indices can be utilized as supporting evidence to interpret judgment agents' trust in automated decision-aids in conjunction with the trust grid discussed previously. Judgments that fall into Quadrant II, and IV in Figure 24 may display these two indices of inter-system conflict and the consequential level of a judge's trust in the automated decision-aid, depending on the level of the coefficient *G*. Again, these can be very useful indices to link human judge's judgment performance and trust in automated decision-aids.

4.0 METHODS

4.1 Introduction

Research findings from the previous studies in human operator's trust in automation were used to generate a set of general research questions. An experiment was designed to investigate the issues raised in chapter 3. This experiment used an aircraft identification task, in which participants were asked to identify unknown objects on a display based on the information provided. There are several reasons why the aircraft identification task was selected. First, the task challenges the human operator's cognitive ability to integrate a set of information to identify the unknown environmental state and compare it to an automated decision-aid. Second, the environment has an uncertain probabilistic nature between the set of information provided to human operator and the environmental state. Third is related to the levels of automation. This task can provide the human operator with the final authority to approve or veto an automatic system based on the previous behavior of the automatic system. An automated system can perform this task independently of the human operator, providing an opportunity to examine the automated system's behavior. Finally, current technological development to integrate the information automatically to support the human operator for judgment and decision-making, called data fusion technology has been proposed for similar tasks.

In this aircraft identification task environment, estimates from a decision-aid were provided in order to characterize the relationship between the decision-aid's characteristics and the operator's trust in the decision-aid. A set of three independent variables was manipulated to control the various aspects of the automated decision-aid and to see their effects on operator's trust in the decision-aid and the level of performance: decision-aid reliability, validity, and understandability.

4.2 Independent Variables

An experimental framework was established and described in the previous chapter. These independent variables have been selected to manipulate various aspects of the automated decision-aids, in order to assess the degree to which they alter a person's reliance on the decision-aid. The independent variables were selected based on the previous research on human operators' trust in automation. In this section, these independent variables will be discussed conceptually, and operationalized in the context of a decision-aid which provides probabilistic estimates regarding a situation assessment.

4.3 Development of Experimental Testbed

A simulated bomb-detection task was selected as a testbed to investigate various issues around human automation interaction in terms of human trust and self-confidence. A Visual Basic® based stand-alone program was written.

Figure 26 shows a 10X10 grid simulated landmine field. The upper right corner, indicated as "chemicals," displays the amount of chemicals the automated decision-aid detected once the operator moves the ADA to a grid. These three chemicals were selected based on the extensive search on possible combinations of important chemicals to make a landmine. For safety purposes, the names of the chemicals were not displayed on the

screen. Also, the color bars located at the top of the display indicate the probability of a landmine on the grid that the ADA has located. The operator was asked to reach the end point as soon as possible without detonating the landmine (wrong decisions).

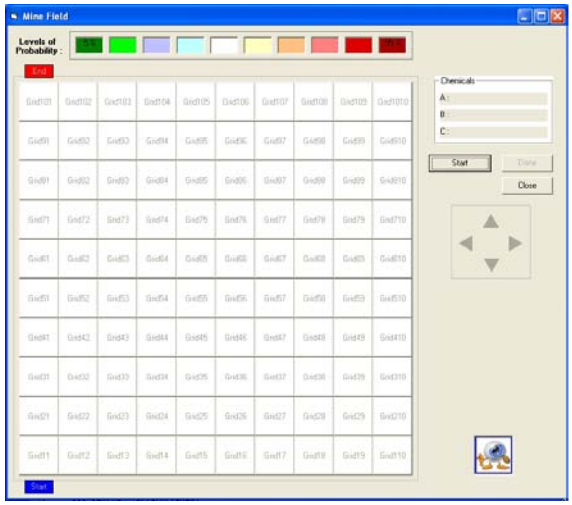

Figure 26: The Initial Screen of the Experimental Testbed Built in this Study

The start (blue box at the bottom) and end points (red box at the top) were randomly selected for each trial. Once the program starts, the operator moves the ADA to the first grid, which displays the levels of chemicals detected in the upper right corner. The operator is pre–trained to understand the relationship between each chemical and the possibility of the presence of a landmine. The operator is asked to make the first decision solely based on the information about the chemical composition. Figure 27 shows this step.

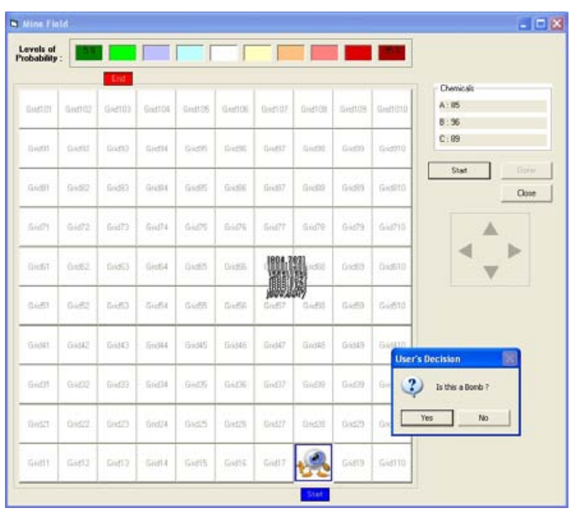

Figure 27: The Second Screen Asks for the Operator's Decision (First Decision) Based on the Detected Chemicals Displayed in the Upper Right Corner

The operator's first decision, shown in Figure 27, brings the next screen, Figure 28. This screen displays the probability of what ADA produces based on the same chemical composition that the operator was already given. When the ADA's estimate is similar to the operator's assessment, the first decision, the operator is asked to accept the ADA's estimate. Otherwise, the operator is asked to reject the ADA's estimate by clicking the "No" button.

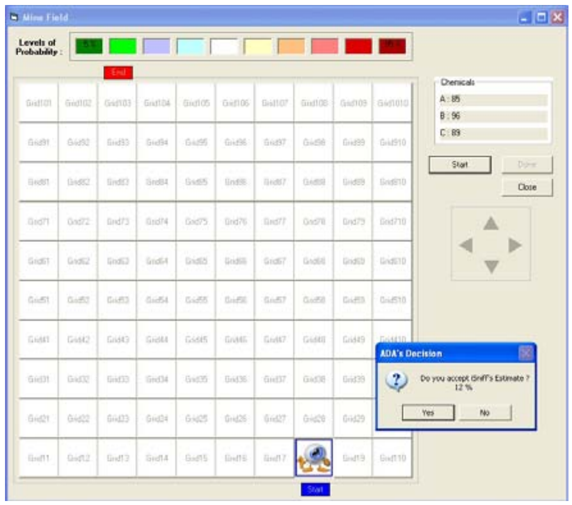

Figure 28: The Third Screen Asks for the Operator's Second Decision after being Informed from the ADA the Probability of Landmines in the Grid

For the final decision, the operator is asked again to make the final on the presence of the landmine on the grid, shown in Figure 29. The previous two steps, or decisions, were designed to investigate any changes that the operator could have made after the operator received the ADA's estimate.

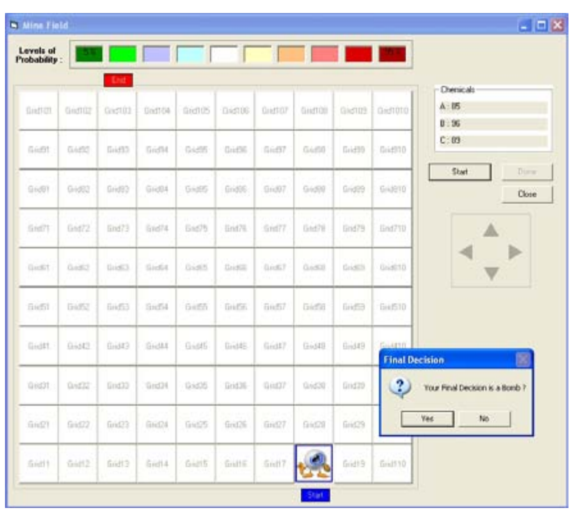

Figure 29: The Fourth Screen Asks for the Operator's Final Decision

Additionally, Figure 30 shows the screen asking the operator's input on the level of self-confidence on the final decision made. This screen is designed to investigate how self-confidence plays a role on making the final decision. In the grid, three letters were displayed along with the ADA's estimate in percentage, i.e., B/N/N 12 percent. The first letter (B in Fig. 27) indicates the operator's first decision (without any input from ADA). The second letter (N) indicates the operator's acceptance of the ADA's estimate (Fig. 28), while the third letter (N) indicates the operator's final decision (Fig. 29). Once the operator finishes the self-confidence rating, the operator is given the feedback that indicates whether the decision was correct or not.

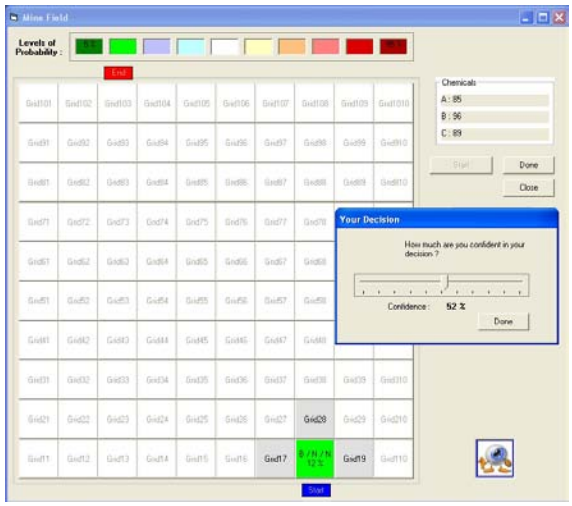

Figure 30: The Fifth Screen Asks for the Operator's Level of Self-confidence on the Final Decision Made

Once the operator finishes with the current grid, potential or adjacent grids that the operator can move within the ADA are highlighted in gray. There are four adjacent grids to any grid in this display (up, bottom, left, and right). However, for the starting point, there are only three adjacent grids highlighted, shown in Figure 31.

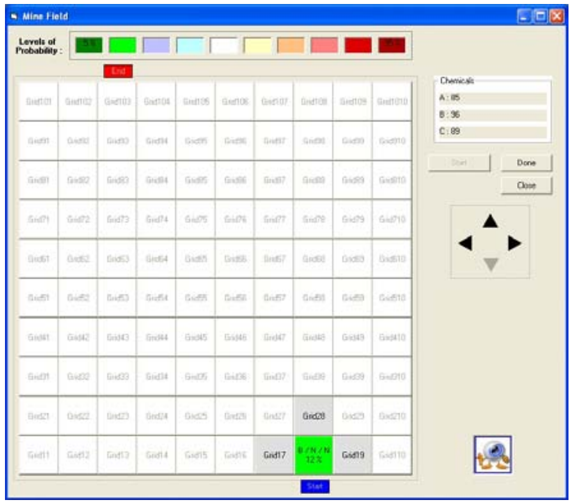

Figure 31: The Next Screen Shows the Possible Grids that the Operator Can Move Once the Operator Made the Final Decision on the Previous Grid (Grayed Grids)

Figure 32 shows the path from the starting to end point. This figure is generated only for demonstration purposes. This screen allows the operator to see how the operator moved through the landmine field in order to develop a better or same (if demonstrating a good level of performance at the first trial) strategy for the upcoming trials.

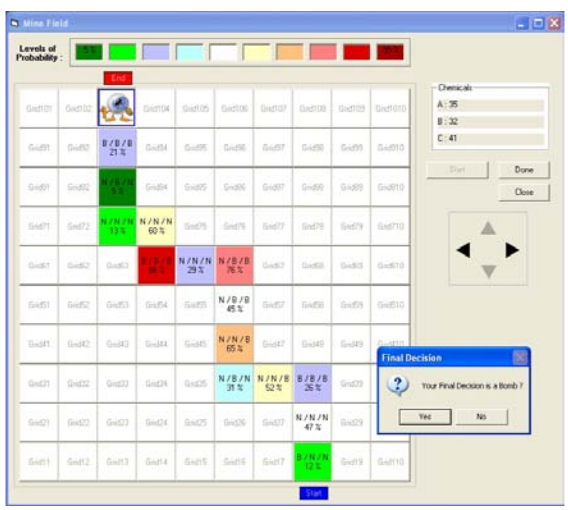

Figure 32: The Screen Showing the Path that the Operator Went Through. Colors Indicate the Level of Probability that the ADA Made Assessments for the Grid (this Figure is for Demonstration Only)

Finally, two screens are given to rate the operator's trust (Fig. 33), and importance of the trust characteristics (Fig. 34). Figure 33 shows the trust questionnaire, which consists of 12 questions (Llinas, Bisantz, Drury, Seong, & Jian,1998). The operator is

asked to rate one's level for each question based on the entire session. This will provide a useful way to investigate the dynamics of operator's trust on the ADA.

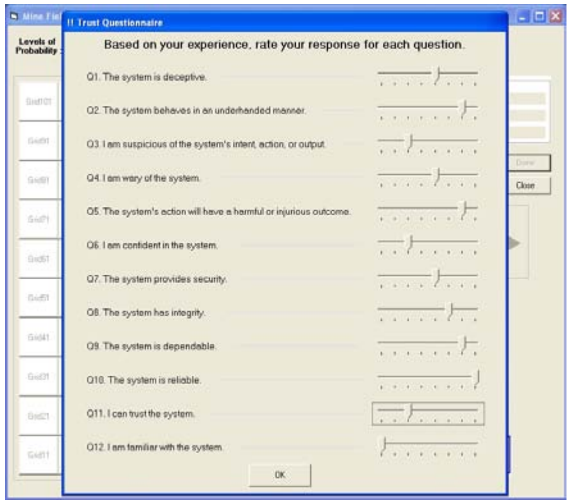

Figure 33: The Trust Questionnaire Screen. The Operator is Asked to Fill Out Twelve Questions

The last screen, Figure 34, asks the operator's input on the rating of importance of the trust characteristics discussed in the previous chapter. This screen is designed to provide a better understanding about the trust continuum for the potential extension of this study.

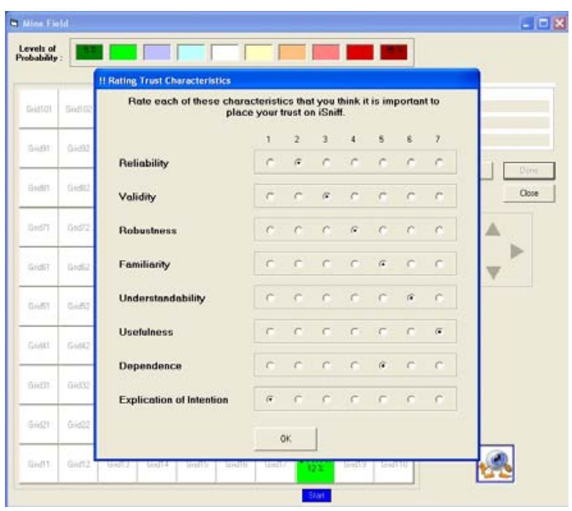

Figure 34: The Final Screen Asking to Rate the Importance of the Characteristics to Produce the Final Level of Trust

5.0 REFERENCES

Balzer, W.K., Doherty, M.E., & O'Connor, R. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin*, 106(3), 410-433.

Balzer, W.K., Sulsky, L.M., Hammer, L.B., Sumner, K.E., Birchenough, T.R., Martens, S.P., & Raymark, P.H. (1992). Task information, cognitive information, or functional validity information: Which components of cognitive feedback affect performance? *Organizational Behavior and Human Decision Processes,* 53, 35-54.

Barber, B. (1983). *The logic and limits of trust*. New Brunswick, NJ: Rutgers University Press.

Beth, T., Vorcherding, M., & Klein, B. (1994). Valuation of trust in open network. In D. Gollman (Ed.), *Computer Security - ESORICS 94, Third European Symposium on Research in Computer Security* (pp. 3-18). Brighton, UK: Springer-Verlag.

Bisantz, A.M. (1997). *Modeling environmental uncertainty to understand and support dynamic decision making* (Ph.D. thesis). Atlanta, GA: Georgia Institute of Technology.

Bisantz, A.M., & Seong, Y. (2001). Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics,* 28, 85-97.

Bowlby, J. (1982). *Attachment and loss. Vol. 1: Attachment* (2nd Ed.). New York: Basic Books.

Brunswik, E. (1952). *The conceptual framework of psychology* (Vol. 1). Chicago: University of Chicago Press.

Bulu, S.T., & Yildirim, Z. (2008). Communication behaviors and trust in collaborative online teams. *Education Technology and Society*, 11(1), 132-147.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.

Christianson, B., & Harbison, W.S. (1997). Why isn't trust transitive? *Proceedings of the International Workshop on Security Protocols* (pp. 171-176).

Cohen, M., Parasuraman, R., & Freeman, J. (1998). Trust in decision aids: A model and its training implications. *Proceedings, 1998 Command and Control Research and Technology Symposium*. Department of Defense C4ISSR Cooperative Research Program.

Cooksey, R.W. (1996). Judgment analysis: Theory, methods, and applications. New York: Academic Press.

Coppola, N., Hiltz, S.R., & Rotter, N. (2001). Building trust in virtual teams. *Proceedings of IEEE Professional Communication Society International Professional Communication Conference* (pp. 353-366). Santa Fe, NM.

Coutu, D.L. (1998). Trust in virtual teams. *Harvard Business Review*, 76(3), 20-21.

Crocoll, W.M., & Coury, B.G. (1990). Status or recommendation: Selecting the type of information for decision aiding. *Proceedings of the Human Factors & Ergonomics Society 34th Annual Meeting* (pp. 1524-1528).

De Vries, M.F.R. (1999). High-performance teams: Lessons from the pygmies. *Organizational Dynamics*, 27, 66-77.

Deutsch, M. (1958). Trust and suspicion. *Journal of Conflict Resolution,* 2, 265-279.

Draper, N.R., & Smith, H. (1966). *Applied regression analysis*. New York: John Wiley & Sons.

Dzindolet, M.T., Pierce, L.G., Beck, H.P., & Dawe, L.A. (1999). Automation reliance on a combat identification system. *Proceedings of the Human Factors & Ergonomics Society 43rd Annual Meeting.*

Endsley, M.R., & Kaber, D.B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics,* 42(3), 462-492.

Endsley, M.R., & Kiris, E.O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors,* 37(2), 381-394.

Feldman, J.M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology*, 66(2), 127-148.

Fitts, P.M. (1951). *Human engineering for an effective air navigation and traffic control system*. Columbus, OH: Ohio State University Research Foundation.

Gallivan, M.J. (2001). Striking a balance between trust and control in a virtual organization: A content analysis of open source software case studies. *Information Systems Journal*, 11, 277-304.

Gattie, G. & Bisantz, A.M. (2002). Utilizing dynamic cognitive feedback to facilitate learning on diagnostic tasks. *Proceedings of the Human Factors & Ergonomics Society 46th Annual Meeting.*

Gefen, D. (2000). E-commerce: the role of familiarity and trust. *Omega,* 28, 725-737.

Glover, S.M., Prawitt, D.F., & Spilker, B.C. (1997). The influence of decision aids on user behavior: Implications for knowledge acquisition and inappropriate reliance. *Organizational Behavior and Human Decision Processes, 72*(2), 232-255.

Grounds, C.B., & Wiley, D. (2001). Alternative methods of mitigating automation distrust impacts. *Proceedings of the Human Factors & Ergonomics Society 45th Annual Meeting* (pp. 1728-1732).

Hamm, R.M. (1988). Moment-by-moment variation in experts' analytic and intuitive cognitive activity. *IEEE Transactions on Systems, Man, and Cybernetics*, 18(5), 757-776.

Hammond, K.R., & Brehmer, B. (1973). Quasi-rationality and distrust: Implications for international conflict. In L. Rappoport & D. Summers (Eds.), *Human judgment and social interactions*. New York: Holt, Rineholt, & Winston.

Hammond, K. R., & Smith, T. H. (1975). *Judgments concerning the future academic goals of the University of Colorado.* Boulder, CO: Center for Research on Training, University of Colorado.

Hammond, K. R., Stewart, T.R., Brehmer, B., & Steinmann, D.O. (1975). Social judgment theory. In M. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes* (pp. 271-312). New York, NY: Academic Press.

Hammond, K. R., & Adelman, L. (1976). Science, values, and human judgments. *Science,* 194, 389-396.

Hammond, K. R., Hamm, R.M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics*, 17(5), 753-770.

Handy, C. (1995). Trust and the virtual organization: How do you manage people whom you do not see? *Harvard Business Review*, 73(3), 40-50.

Hanes, R.M., & Gebhard, J.W. (1966). Information requirements for the control of combat forces. In J. S. Albanes (Ed.), *Reports of the committee on vision, 1947-1990*. Washington, DC: National Academy Press.

Hopkins, V.D. (1994). Human performance implications of air-traffic control automation. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Recent research and trends* (pp. 314-319). Mahwah, NJ: Lawrence Erlbaum.

Hursch, C.J., Hammond, K.R., & Hursch, J.L. (1964). Some methodological considerations in multiple-cue probability studies. *Psychological Bulletin*, 71, 42-60.

Iacono, C.S., & Weisband, S. (1997). Developing trust in virtual teams. *Proceedings of the 30th Annual Hawaii International Conference on System Sciences* (Vol. 2: 412-420).

Ishaya, T., & Macaulay, L. (1999). The role of trust in virtual teams. *Electronic Journal of Organizational Virtualness,* 1(1), 140-157.

Itoh, M., & Tanaka, K. (2000). Mathematical modeling of trust in automation: Trust, distrust, and mistrust. *Proceedings of the IEA2000/HFES2000 Congress* (pp. 1-9). San Diego, CA.

Jarvenpaa, S.L., Knoll, K., & Leidner, D.E., 1998. Is anybody out there? The implications of trust in global virtual teams. *Journal of Management Information Systems*, 14(4), 29-64.

Jarvenpaa, S.L., & Leidner, D.E. (1999). Communication and trust in global virtual teams. *Organization Science,* 10(6), 791-865.

Jian, J.Y., Bisantz, A.M., & Drury, C.G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 1(4), 53-71.

Jones, S., & Marsh, S. (1997). Human-computer-human interaction: Trust in CSCW. *ACM SIGCHI Bulletin,* 29(3), 36-40.

Kanawattanachai, P., & Yoo, Y. (2002). Dynamic nature of trust in virtual teams. *Journal of Strategic Information Systems*. 11, 187-213.

Kim, J., & Moon, J.Y. (1998). Designing towards emotional usability in customer interfaces - trustworthiness of cyber-banking system interfaces. *Interacting with Computers*, 10(1), 1-29.

Kirlik, A. (1993). Modeling strategic behavior in human-automation interaction: Why an "Aid" can (and should) go unused. *Human Factors,* 35(2), 221-242.

Larzelere, R.E., & Huston, T.L. (1980). The dyadic trust scale: toward understanding interpersonal trust in close relationships. *Journal of Marriage and the Family*, 42, 595-604.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocations of functions in human-machine systems. *Ergonomics,* 35(10), 1243-1270.

Lee, J., & Moray, N. (1994). Trust, self-confidence and operators' adaptation on automation. *International Journal of Human-Computer Studies,* 40, 153-184.

Lee, J.D., Gore, B.F., & Campbell, J. L. (1999). Display alternatives for in-vehicle warning and sign information: Message style, location, and modality. *Transportation Human Factors Journal*, 1(4), 347-377.

Lerch, F.J., & Prietula, M.J. (1989). How do we trust machine advice? In G. Salvendy & M.J. Smith (Eds.), *Designing and using human-computer interface and knowledge based systems* (pp. 410-419). Amsterdam: Elsevier Science Publishers.

Lewis, J.D., & Weigert, A.J. (1985). Trust as a social reality. *Social Forces*, 63, 967-985.

Libby, R. & Luft, J. (1993). Determinants of judgment performance in accounting settings: Ability, knowledge, motivation, and environment. *Accounting, Organizations, and Society*, 18(5), 425-451.

Llinas, J., Bisantz, A., Drury, C.G., Seong, Y., & Jian, J. (1998). *Studies and analyses of aided adversarial decision-making, Phase 2: Research on human trust in automation.* New York: Center for Multisource Information Fusion, State University of New York, Buffalo.

McAllister, D.J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38(1), 24-59.

Meyerson, D., Weick, K.E., & Kramer, R.M. (1996). Swift trust and temporary groups. In R.M. Kramer & T.R. Tyler (Eds.), *Trust in organizations, frontiers of theory and research* (pp. 166-195). Newbury Park, CA: Sage Publications.

Mockler, R.J. (1989). *Knowledge-based systems for management decisions*. New Jersey: Prentice Hall.

Moffa, A.J., & Stokes, A.F. (1996). Trust in a medical system: can we generalize between domains? In M. Mouloua & J.M. Koonce (Eds.), *Human-automation interaction: Research and practice* (pp. 218-224). Cocoa Beach, FL: CRC Press.

Moon, Y., & Nass, C. (1998). Are computers scapegoats? Attributions of responsibility in human-computer interaction. *International Journal of Human Computer Studies,* 49, 79-94.

Moray, N. (1997). Human factors in process control. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 1944-1971). New York: Wiley.

Morphew, M.E., & Wickens, C.D. (1998). *Pilot performance and workload using traffic displays to support free flight. Proceedings of the Human Factors & Ergonomics Society 42nd Annual Meeting.*

Morris, S.A., Marshall, T.E., & Rainer, R.K. (2002). Impact of user satisfaction and trust on virtual team members. *Information Resources Management Journal*, 15(2), 22-30.

Morrison, J.G., Kelly, R.T., Moore, R.A., & Hutchins, S.G. (1998). Implications for decision-making research for decision support and displays. In J.A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 375-406). Washington, DC: American Psychological Association.

Mosier, K.L., & Skitka, L.J. (1996). Human decision makers and automated decision aids: Made for each other. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and application* (pp. 201-220). Mahwah, NJ: Lawrence Erlbaum.

Muir, B.M. (1987). Trust between human and machines, and the design of decision aids. *International Journal of Man-Machine Studies,* 24, 527-539.

Muir, B.M., & Moray, N. (1987). Operator's trust in relations to system faults. *IEEE International Conference on Systems, Man, and Cybernetics* (pp. 258-263). Alexandria, VA.

Muir, B.M. (1994). Trust in automation, Part I: Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics,* 37(11), 1905-1922.

Muir, B.M., & Moray, N. (1996). Trust in automation, Part II: Experimental studies of trust and human intervention in a process control simulation. *Ergonomics,* 39(3), 429-460.

Norman, D.A. (1990). The "problem" of automation: Inappropriate feedback and interaction, not "over-automation". *Philosophical Transactions of the Royal Society of London,* B327, 585-593.

Parasuraman, R., Molloy, R., & Singh, I.L. (1993). Performance consequences of automation-induced "complacency". *International Journal of Aviation Psychology,* 3, 1-23.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics,* 30(3), 286-297.

Powell, A., Piccoli, G., & Ives, B. (2004). Virtual teams: A review of current literature and direction for future research. *The DATA BASE for Advances in Information Systems*, 35(1), 6-36.

Rasmussen, J. (1983). Skills, rules, knowledge: Signal, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, 13, 257-266.

Rasmussen, J., Pejtersen, A.M., & Goodstein, L.P. (1994). *Cognitive systems engineering*. New York: John Wiley & Sons.

Reason, J. (1990). *Human error*. Cambridge, UK: Cambridge University Press.

Rempel, J.K., Holmes, J.G., & Zanna, M.P. (1985). Trust in close relationship. *Journal of Personality and Social Psychology, 49*(1), 95-112.

Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and application* (pp. 22-27). Mahwah, NJ: Lawrence Erlbaum.

Rotter, J.B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35, 651-665.

Sarker, S., Lau, F., & Sahay, S. (2001). Using an adapted grounded theory approach for inductive theory building about virtual team development. *The DATA BASE for Advances in Information Systems*, 32(1), 38-56.

Sarker, S., Valacich, J.S., & Sarker, S. (2003). Virtual team trust: Instrument development and validation in an IS educational environment. *Information Resources Management Journal*. 16(2), 35-55.

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors,* 37(1), 5-20.

Seong, Y., & Bisantz, A. (2008). The impact of cognitive feedback on judgment performance and trust with decision aids. *International Journal of Industrial Ergonomic,* 38, 608-625.

Sheridan, T.B., & Verplank, W.L. (1978). *Human and computer control of undersea teleoperators*. Cambridge, MA: MIT Man-Machine Systems Laboratory.

Sheridan, T.B., & Hennessy, R.T. (1984). *Research and modeling of supervisory control behavior*. Washington, DC: National Academic Press.

Sheridan, T.B. (1988). Trustworthiness of command and control systems. *Proceedings of IFAC Man-Machine Systems* (pp. 427-431). Oulu, Finland.

Sheridan, T.B. (1992). Telerobotics, automation, and human supervisory control. Cambridge, MA: MIT Press.

Thomas, D.M., & Bostrom, R.P. (2008). Building trust and cooperation through technology adaptation in virtual teams: Empirical field evidence. *Proceedings of the 41st Hawaii International Conference on System Sciences* (pp. 423-433).

Toulmin, S. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.

Tucker, L.R. (1964). A suggested alternative formulation in the developments of Hursch, Hammond, & Hursch and Hursch, Hammond, & Todd. *Psychological Review*, 71(6), 528-530.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science,* 185, 1124-1131.

Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics,* 22, 589-606.

Wickens, C.D. (1992). *Engineering psychology and human performance* (2nd Ed.). New York: Harper Collins.

Wiener, E.L., & Curry, R.E. (1980). Flight-deck automation: Promises and problems. *Ergonomics,* 23, 995-1011.

Wiener, E.L. (1988). Cockpit automation. In E.L. Wiener & D.C. Nagel (Eds.), *Human factors in aviation*. San Diego, CA: Academic Press.

Will, R.P. (1991). True and false dependence on technology: Evaluation with an expert system. *Computers in Human Behavior,* 7, 171-183.

Woods, D.D. (1988). Coping with complexities: The psychology of human behavior in complex system. In L.P. Goodstein, H.B. Anderson & S.E. Olsen (Eds.), *Tasks, errors and mental models* (pp. 128-148). New York: Taylor and Francis.

Woods, D.D. (1993). Price of flexibility in intelligent interfaces. *Knowledge-Based Systems*, 6(4), 189-196.

Woods, D.D. (1996). Decomposing automation: Apparent simplicity, real complexity. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and application* (pp. 3-17). Mahwah, NJ: Lawrence Erlbaum.

Yeh, M., & Wickens, C.D. (2001). Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. *Human Factors*, 43, 355-365.

Zeide, J.S., & Liebowitz, J. (1987). Using expert systems: The legal perspective. *IEEE Expert,* 2(1), 19-22.

APPENDIX A

Literature Reviewed with Human Trust in Virtual Teams

Although literature with regards to the issue of human trust in virtual environments is limited, the list below provides articles and papers that were reviewed.

Abdul-Rahman, A., & Hailes, S. (2000). Supporting trust in virtual communities. *Proceedings of the 33rd Hawaii International Conference on System Sciences* (Volume 6, page 6007).

Aubert, B.A., & Kelsey, B.L. (2003). Further understanding of trust and performance in virtual teams. *Small Group Research*, 34(5), 575-618.

- Based on the theoretical frameworks by Mayer and McAllister.
- Mayer: perceived ability, integrity, benevolence, and propensity to trust.
- McAllister: cognitive and affective trust.
- Investigated the relationships between the antecedents of trust, trust formation, and performance outcomes.
- H1a – partially supported: perceived I & A were significant for both groups, but not B.
- H1b – not clear support: propensity to trust was only significant for final/remote evaluation.
- H1c – not clear: not consistent results across the different levels of propensity groups.
- H2 – approved: higher perceived integrity and benevolence for local than remote. (But never changed over time. – H3a &b).
- H3a and H3b – disapproved: virtual teams failed to establish the conditions necessary for trust to develop effectively during the process.
- H4 – disapproved: initially cognitive trust was higher for local than remote. The gap between two groups increased over time (local increasing trust, while remote decreasing trust over time – although not significant).
- H5 – disapproved: performance – the level of trust among teammates has a negligible effect on team performance. But the final levels of trust between the local and remote subteams did have a significant impact on effective performance.
- Additional idiosyncratic analyses were provided.

Bulu, S.T., & Yildirim, Z. (2008). Communication behaviors and trust in collaborative online teams. *Education Technology and Society*, 11(1), 132-147.

- Investigated the relationship between collaborative communication behaviors (leadership, feedback, social interaction, enthusiasm, task and technical uncertainties, and task-oriented interactions) in the online learning environment.
- Results showed that trust can be developed in an online environment (contrary to others who demonstrated that trust needs face-to-face interaction).
- Different trust levels showed different collaborative communication behaviors, which is consistent with other findings (p142).
- There is a division of labor and equal participation in the group work, the reasons for conflicts and disagreements are minimized, and trust deepens.
- Regular communication is necessary to deepen trust.

- Higher trust groups took initiatives and there were roles for all, while lower trust groups showed unequal distributed communications. Might be due to ineffective and negative behaviors of the leader. The presence of additional notices to urge others to contribute to a group effort.
- More positive and encouraging feedback in the higher trust groups to maintain social presence, and climate in the group.
- Uncertainty statements decreased in the higher trust groups, which could be the results of member support and social interaction in the groups.
- Initial enthusiasm and social interaction help build trust in the early period, which might result from learning styles, socio-emotional needs, task content in addition to the social interaction, and expression of their preferences.

Coppola, N.W., Hiltz, S.R., & Rotter, N.G. (2004). Building trust in virtual teams. *IEEE Transactions on Professional Communication*, 47(2), 95-104.

Furst, S., Blackburn, R., & Rosen, B. (1999). Virtual team effectiveness: A proposed research agenda. *Information Systems Journal*, 9, 249-269.

Gibson, C.B., & Cohen, S.G. (2003). *Virtual teams that work: Creating conditions for virtual team*. San Francisco, CA: Jossey-Bass.

Greenburg, P.S. (2006). *Mastering virtual teams: Strategies tools, and techniques that succeed* (3rd Ed). San Francisco, CA: Jossey-Bass.

Holton, J.A. (2001). Building trust and collaboration in a virtual team. *Team Performance Management*, 7(3/4), 36-47.

Johnson, S.D., Suriya, C., Yoon, S.W., Berrett, J.V., & Fleur, J.L. (2002). Team development and group processes of virtual learning teams. *Computers & Educations*, 39, 379-393.

Lipnack, J., & Stamps, J. (1999). The new way to work. *IEEE Engineering Management Review*, Winter, 90-95.

Montoya-Weiss, M.M., Masey, A.P., & Song, M. (2001). Getting it together: Temporal coordination and conflict management in global virtual teams. *The Academy of Management Journal*, 44(6), 1251-1262.

Ngo-Mai, S., & Raybaut, A. (2007). *Swift trust and virtual team dynamics*. Valbonne, France: Universite de Nice-Sohpia Antipolis/CNRS DEMOS-GREDEG.

Pauleen, D., & Yoong, P. (2001). Facilitating virtual team relationship via internet and conventional communication channels. *Internet Research: Electronic Networking Applications and Policies*, 11(3), 190-202.

Ridings, C.M., Gefen, D., & Arinze, B. (2002). Some antecedents and effects of trust in virtual communities. *Journal of Strategic Information Systems*, 11, 271-295.

Sarker, S., & Sahay, S. (2002). Understanding virtual team development: An interpretive study. *Journal of the Association for Information Systems*, 3, 247-285.

Tung, L., & Turban, E. (1998). A proposed research framework for distributed group support systems. *Decision Support Systems*, 23, 175-188.

Warkentin, M., & Beranek, P.M. (1999). Training to improve team communication. *Information Systems Journal*, 9, 271-289.

Williamson, O.E. (1993). Calculativeness, trust, and economic organizations. *Journal of Law and Economics*, 36(1), 453-486.

APPENDIX B

Literature Reviewed with Human Trust in Sociological Trust

Lerch and Prietula (1989): Source pedigree; to see whether sociological definitions of trust can be extended to the engineering field, varied in human expert, human novice, and expert system. Confidence was measured to indicate the level of dependability, while agreement with advice was used for a measure of predictability. Eighty-five students were given different sources of advice with varying degrees of extensiveness. That is, they were sometimes given correct information, while in another case given incorrect but little information, and in the last case complicated but correct information. The level of trust was highest with advice given by human experts. There was no difference between human novice and expert system. They raised a somewhat different perspective from Muir's; while Muir suggested that low level of trust associated with more intense activity of monitoring, overestimation of representativeness of small samples can cause overestimation of predictability, which consequently may cause increase in trust. These two perspectives may not be completely different, rather they do not agree on the causality direction. It definitely needs further speculation. "Inertia" was also confirmed as was validated in Muir and Lee and Moray. The authors' accounts of trust characteristics seems somewhat similar to Hammond's theory of cognitive continuum, in that predictability seems to reside in the analytical cognitive side, dependability in the quasi-rational, and faith in the intuitive side.

Felkey, Monk, and Stec (1984): Effect of jamming or deception in the communication network was examined empirically. Jamming the communication line connecting the lowest (radar) to the highest (decision-maker) produced further decrease in performance. The worst performance of all the conditions was caused with jamming the channel from bottom to the top, along with deceiving the channel between the top two sources.

Human Factors Trust

Theoretical Trust

Bainbridge, L. (1983). Ironies of automation. *International Federation of Automatic Control*, 19, 775-779.

Wickens, C.D. (1994). Designing for situation awareness and trust in automation. *IFAC Conference on Integrated Systems Engineering* (pp. 365-370). Baden-Baden, Germany.

Sheridan, T.B. (1980). Computer control and human alienation. *Technology Review*, 10, 61-73.

Sheridan, T.B., Vamos, T., & Aida, S. (1983). Adapting automation to man, culture, and society. *Automatica*, 19, 605-612.

Jian, J., Bisantz, A.M., Drury, C.G., & Llinas, J (1998). *Foundations for an empirically determined scale of trust in automated systems*. Center for Multisource Information Fusion, SUNY Buffalo. CMIF-1-98.

Hwang, P., & Burgers, W.P. (1997). Properties of trust. *Organizational Behavior and Human Decision Processes*, 69, 67-73.

Malone, T.B. (1990). Human factors and human error. *Proceedings of the Human Factors Society 34th Annual Meeting*, 651-654.

Dassonville, I., Jolly, D., & Desodt, A.M. (1994). Towards a modelization of trust in a teleoperation system. In P.T. Kidd & W. Karwowski (Eds), *Advances in agile manufacturing: Integrating technology, organization and people* (pp. 507-510). Macclesfield, UK: Cheshire Henbury.

Westin, A.F. (1992). Two key factors that belong in a macroergonomic analysis of electronic monitoring: Employee perceptions of fairness and the climate of organizational trust or distrust. *Applied Ergonomics*, 23(1), 35-42.

Crocoll, W.M., Ellis, N.C., & Simmons, D.B. (1996). The trust relationship between users and intelligent support systems. IS*AI/IFIS*, 77-84.

Slovic, P. (1993). Perceived risk, trust and democracy. *Risk Analysis*, 13, 675-682.

Dzindolet, M.T., Pierce, L.G., Beck, H.P., & Dawe, L.A. (1999). Misuse and disuse of automation aids. *Proceedings of the Human Factors Society 43th Annual Meeting*, 337-343.

Empirical Trust Research

Muir (1987, 1994a): The first paper that raised and discussed the importance of internal construct trust in automation environment. Followed by two sociological definitions of trust developed by Barber and Rempel et al. Muir developed a framework of trust in human-machine relationships. The meaning of trust is accessed based on Barber's interpretation of trust with expectation (persistence, technically competence, and fiduciary responsibility) while the dynamic characteristic of trust is approached using the Rempel et al., stage model of trust (predictability, dependability, and faith). Also, worst-case scenario training was suggested to improve users' understanding of the automated systems' responsibility and competence.

Muir (1994b): Two experiments were conducted to investigate the nature and dynamic of trust in machines and how trust has an effect on human behavior of intervention on automation. In Experiment 1, six students performed a milk pasteurization plant medium-fidelity simulation, where they had to control two subsystems, a pump, and a heating system. Each subsystem had two failure modes, control and display error, in that each failure is manipulated separately. The level of trust on overall turned out to be best described by competence and then little effect of

responsibility. Further, the order of dynamics of trust development was not validated by the experiment. In fact, the order was in complete reverse in that faith was the better predictor in the early training stage, then dependability, and then finally predictability. Although participants' subjective ratings of trust changed according to the quality of automation, two failure modes had no effect on either participants' performance, or control action behavior or use of automation. An interesting aspect was overlooked in the experiment. Command input (control function) was altered in three different levels (exact, variable, and constant) while display output (display function) was also manipulated in three levels (honest, variable, and constant), to produce the final nine experimental conditions. Among the conditions, two conditions seem identical; exact control manipulation with constant error in display (C2), and constant error in control manipulation with honest display (C4). The levels of trust in control function and in display function were well calibrated according to the true levels of both manipulations in the exact control with constant display manipulation (cell C2). That is, participants' levels of trust changed accordingly when only the display function was manipulated with control function intact. The level of trust in control was rated high, as it should have been, while the level of trust in display was rated low, as it should have been. When the control function was manipulated with display function intact (C4), however, both the level of trust in display function and the level of trust in control function was rated low, which means that the participants could not identify the source of error (presumably). Thus, the effect of low level of trust in control seems to spread over the level of trust in display. Experiment 2 was conducted to see the differential effects of consistent vs. variable errors. Error magnitude has a differential effect on the level of trust as the error magnitude increases, the level of trust decreases in a decreasing rate. Also, the relationship between the level of trust and the use of automation was supported. Monitoring activity was also associated with the level of trust in that fewer operators trusted the system; the more intensely they monitored the system.

Lee and Moray (1992): Four dimensions of trust: foundation, purpose, process, and performance. Total of 19 students performed medium-fidelity simulation of orange juice production plant, which consisted of two controllable subsystems, pump, and heating. Sizes of fault, along with continuity of faults where they are transient or continuous were independently manipulated. With transient fault, the level of trust was changed approximately proportional to fault magnitude with little after-effect, but still hard to recover with the greater fault magnitude. Fault magnitude had an effect on trust, but no effect on performance; thus the size was matter to trust, but not to performance. Chronic fault showed the same pattern with the transient faults. Not able to derive the causality direction between trust and system performance. Contradicting with Muir's result; more use of automation with trust declined.

Lee and Moray (1994): Introduced another factor, self-confidence related with trust to describe use of automation. Use of automation: when trust exceeds self-confidence, operators will revert to manual automation. Individual difference was detected for use of automation. Use of multitrait and multimethod: to identify the relationship between use of automation and trust with self-confidence (convergent and discriminant validity). Order of introduction of automatic vs. manual mode (MA vs. AM) with continuous faults; performance was worse in MA order than in AM order.

Riley (1989): Proposed a causality model among many factors. Also, suggested how trust and compliance in automation should be changed according to machine and system accuracy.

Riley (1996): Conducted 4 experiments to validate his causality model. In Experiment 1 (30 students), automation reliability, workload, and uncertainty were controlled to measure the time of automation use. The results showed that participants indicated preference using manual control over automation, which is the same result as Lee's. Also, they did not hesitate to turn back the automation after automation failure, which is a contradicting result from what Muir or Lee suggested. Experiment 1 also found significant effects of uncertainty and reliability, but not workload. Experiment 2 (three conditions) investigated the effects of the amount of information given (state uncertainty) and trust on automation use. Results showed that both state uncertainty and trust, affect automation use decision, but only early in the subjects' experience with the automation. In Experiment 3 (34 pilots) showed substantial preference to automated control over manual. Uncertainty and reliability was significant but not workload, which is the same result as with the students. Pilots preferred errors of commissions over errors of omission. Experiment 4 was same as the third with 31 pilots, to see the use of automation in relation to their own level of manual proficiency. No relationship was found and there were no differences in CPRS between students and pilots.

Gardner, Dukes, and Discenza (1993): Causal analysis was conducted to investigate the effect of various constructs as self-confidence and attitudes on the computer use. The factors analyzed were computer experience, programming experience, frequency of use, school uses, and confidence in school settings (723 high school students). As they gained experience with computers, the level of anxiety decreased and the confidence level increased, which consequently causes the development of favorable attitudes toward computers. However, direct relationship between computer use and computer attitude was negative, which means that the greater use of computer resulted in negative attitudes toward computers, positive attitudes overall though.

Will (1991): Seven novices and five experts were presented with erroneous expert systems for petroleum, with interpretation tasks, to investigate the degree of dependence on expert systems, and found that the experts did not identify the erroneous system.

Cohen, Parasuraman, and Freeman (1997): A qualitative model of trust in decision-aids was proposed using Toulmin's theory of argument.

Dzindolet, Pierce, Beck and Dawe (1999): A general model of automation use is proposed to identify potential factors affecting the operators' proper use (misuse and disuse) of automation. The model describes three distinct approaches to automation use: cognitive, social, and motivational processes. The cognitive processes refer to the situations where human operators blindly rely on the decisions made by the automated decision-aids instead of going through effortful cognitive processes, thus induced to the automation bias.

Hwang (1997): Built a simple analytical model of trust and proposed seven properties of trust in relation with cooperation, fear, and greed.

Stewart, Heideman, Moninger, and Reagan-Cirincione (1992): Effect of improved information on three Lens Model Equation components in weather forecasting environment was investigated.

Data Fusion Technology

Nigay, L., & Coutaz, J. (1993). A design space for multimodal systems: concurrent processing and data fusion. *Proceedings of INTERCHI*, 172-178. ACM Press.

Svenmarck, P. (2000). Human engineering of multisensor and multisource systems. *Proceedings of SPIE,* 4052, 67-76.

Wohl, J.G., Entin, E. E., Kleinman, D.L., & Pattipati, K. (1984). Human decision processes in military command and control. *Advances in Man-machine Systems Research*, 1, 261-307.

Shahbazian, E., Duquet, J., & Simard, M. (1999). Literature survey on computer-based decision support for command and control systems. *ISIF/FUSION 99*, 926-933.

Paradis, S., Breton, R., & Roy, J. (1999). Data fusion support of dynamic human decision making. *ISIF/FUSION 99*, 445-454.

Other Factors

Pulford, B.D., & Colman, A.M. (1996). Overconfidence, base rates, and outcome positivity/negativity of predicted events. *British Journal of Psychology*, 87, 431-445.

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: it depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes*, 79(3), 216-247.

Milewski, A., & Lewis, S.H. (1997). Delegating to software agents. *International Journal of Human-Computer Studies*, 46, 485-500.

Briggs, P., Burford, B., & Dracup, C. (1998). Modelling self-confidence in users of a computer-based system showing unrepresentative design. *International Journal of Human-Computer Studies,* 49, 717-742.

Soll, J.B. (1996). Determinants of overconfidence and miscalibration: the role of random error and ecological structure. *Organizational Behavior and Human Decision Processes*, 65(2), 117-137.

Gardner, D.G., Dukes, R.L., & Discenza, R. (1993). Computer use, self-confidence, and attitudes: a causal analysis. *Computers in Human Behavior*, 9, 427-440.

General Automation – Methodology

Hammond, K.R., Hamm, R.M., & Grassia, J. (1986). Generalizing over conditions by combining the multitrait-multimethods matrix and the representative design of experiment. *Psychological Bulletin*, 100(2), 257-269.

Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental process. *Psychological Review*, 84(3), 231-259.

Vicente, K.J., & Torenvliet, G.L. (2000). The earth is spherical (p<0.05): alternative methods of statistical inference. *Theoretical Issues in Ergonomics Science*, 1(3,) 248-271.

Hammond, K.R. (1986). Generalization in operational contexts: what does it mean? Can it be done? *IEEE Transactions on SMC*, 16, 428-433.

Stewart, T.R., Middleton, P., Downton, M., & Ely, D. (1984). Judgments of photographs vs. field observations in studies of perception and judgment of the visual environment. *Journal of Environmental Psychology*, 4, 283-302.

Booske, B.C., & Sainfort, F. (1998). Relation between quantitative and qualitative measures of information use. *International Journal of Human-Computer Interaction*, 10, 1-21.

Kelley, C.L., Bosman, E.A., Charness, N.H., & Mottram, M.A. (1998). Reliability assessment of a user proficiency measurement technique. *International Journal of Human-Computer Interaction*, 10, 33-49.

Jones, P.M., Chu, R.W., & Mitchell, C.M. (1995). A methodology for human-machine systems research: knowledge engineering, modeling and simulation. *IEEE Transactions on SMC*, 25, 1025-1038.

Cognitive Continuum

Hammond, K.R., Hamm, R.M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on SMC*, 17(5), 753-770.

Hamm, R.M. (1988). Moment-moment variation in experts' analytic and intuitive cognitive activity. *IEEE Transactions on SMC*, 18(5), 757-776.

Unreliable Information

Brehmer, B. (1970). Inference behavior in a situation where the cues are not reliably perceived. *Organizational Behavior and Human Decision Processes*, 5, 330-347.
Brehmer investigated human inference behavior when the given information was not reliably perceived. Here, performance variation was not caused or manipulated by any experimental design methods. Rather by giving natural judgmental environments to the judges, unreliably perceived information, so called self-generated unreliability, was blamed on the judges' non-perfect performance. He concluded that unreliability in the perception of the cue values influenced the judges' performance. Specifically, unreliability in the perception was reflected in the validity of the information available for the inference. In these experiments, however, validity of information, cue validity in Brunswik's term, was confounded with task variation, which was perceived as the lack of reliability in the perception. As Brehmer pointed out, the uncertainty lays in between the physical cue values and the criterion values. Thus, distinguishing cue validity with cue reliability was precluded.

O'Connor, R.M. Jr., Doherty, M.E., & Tweney, R.D. (1989). The effects of system failure error on predictions. *Organizational Behavior and Human Decision Processes*, 44, 1-11.

York, K.M., Doherty, M.E., & Kamouri, J. (1987). The influence of cue reliability on judgment in a multiple cue probability learning task. *Organizational Behavior and Human Decision Processes*, 39, 303-317.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.

Birnbaum, M.H., Wong, R., & Wong, L. K. (1976). Combining information from sources that vary in credibility. *Memory & Cognition*, 4, 330-336.

Gaeth, G.J., & Shanteau, J. (1984). Reducing the influence of irrelevant information on experienced decision makers. *Organizational Behavior and Human Decision Processes*, 33, 263-282.

Doherty, M.E., & Sullivan, J.A. (1989). $\varrho = \boldsymbol{p}^{\underline{*1}}$ **.** *Organizational Behavior and Human Decision Processes*, 43(1), 136-144.

St. John, M., Oonk, H.M., & Osga, G.A. (2000). Designing displays for command and control supervision: Contextualizing alerts and "trust-but-verify" automation. In *Proceedings of the IEA2000/HFES 200 Congress* (pp. 646-649).

Halpin, S.M., Johnson, E.M., & Thornberry, J.A. (1973). Cognitive reliability in manned systems. *IEEE Transactions on Reliability*, 22, 165-169.

Felkey, M.A., Monk, D L., & Stec, L.J. (1984). *The effect of jamming/deception on decision making*. AFAMRL-TR-84-055 (ADA149740).

Barkan, R., Zohar, D., & Erev, I. (1998). Accidents and decision making under uncertainty: a comparison of four models, *Organizational Behavior and Human Decision Processes*, 74, 118-144.

Bragger, J.D., Bragger, D., Hantula, D.A., & Kirnan, J. (1998). Hysteresis and uncertainty: the effect of uncertainty on delays to exit decisions. *Organizational Behavior and Human Decision Processes*, 74, 229-253.

Gorman, M.E. (1986). How the possibility of error affects falsification on a task that models scientific problem solving. *British Journal of Psychology*. 77, 85-96.

APPENDIX C

Literature Reviewed with Human Trust in a Varity of Environments

Title: Modeling the impact of trust and distrust in agent networks
Authors: G. Gans, M. Jarke, S. Kethers, G. Lakemayer
Summary: The main goal of this paper is to develop a methodology, based on the TCD (Trust-Confidence-Distrust) model, and is suitable for the specification and analysis of social networks. The researcher focused on formalizing trust in social networks, a recent form of "Coopetition," that promises to combine the benefit of two traditional coordination mechanisms of modern societies. The researchers motivate and develop their TCD model. The researchers of this paper have also agreed with previous researchers about trust modeling of goals and dependency is crucial to understand networks in general and in particular, the role of trust confidence, and distrust within this networks. Finally, although coordination by means of trust and confidence can enable and facilitate cooperation, it has its cost, in networks; trust and confidence need to be watchful. The TCD model requires an extended modeling formalism, which does not restrict itself to structural aspect of trust.
Source: http://www-kbsg.informatik.rwth-aachen.de/literature/Gans2001Modeling.pdf

Title: A model for predicting human trust in automated systems
Authors: Mohammad T. Khasawneh, Shannon R. Bowling, Xiaochun Jiang, Anand K. Gramopadhye, Brian J. Melloy
Summary: Trust is an important intervening variable between an automated system and its use and consequent performance. Current approaches to trust measurement rely solely on subjective metrics. Moreover, existing models are based on different theoretical concepts of trust between humans that may not necessarily be as relevant to machines. More importantly, researchers have not been able to link system performance and trust using a precise quantitative model. The results revealed that human trust is directly related to changes in the performance of the computerized system. Therefore, the approach followed in this study illustrates that trust can be mathematically predicted by knowing system errors, an objective approach, which provides a more accurate measurement of human trust. This is a very important finding suggesting that questionnaires based on dimension of trust, are not the best method for measuring this construct, as if dimensions had the same impact on users. However, this study is only the first step and further research is necessary in order to better understand the relationship between human trust and system performance. The results of this research addressing the issue of trust in human-machine system have obvious implications for both the theoretical human factors researcher and the practitioner.
Source:
http://www.ws.binghamton.edu/mkhasawn/Publications/Conference/IJIE/IJIE%202003%20(Trust).pdf

Title: The trust relationship between users and intelligent support systems
Authors: William M. Crocoll, Newton C. Ellis, Dick B. Simmons
Summary: The purpose of the paper is to study and combine/compare human and machine trust theory from five different previous researchers who have studied trust modeling use. There are essentially three views on the relationship between the intelligent machine and the human. One view is that the decision-aid should act on the cognitive prosthesis that compensates for deficiencies in the human's abilities. Another view is that the decision-aid is a cognitive instrument/tool in the hands of a competent, but resource limited to the decision-maker. A third view suggests that both forms of assistance be presented to provide redundancy. The calibration of trust between human and machines are not equally competent and certainly not flawless, thus the user must learn to calibrate trust to machine trust worthiness. The extreme of mistrust is trusting an incompetent machine. The human may place too much trust in the intelligent system endowing it with almost mystical abilities. On the other hand, complete distrust of a competent machine may result in the user doing all of the work leading to overload and decrease in overall system performance.
Source:
http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/6956/18719/00864103.pdf?arnumber=864103

Title: The effect of external safeguards on human-information system trust in an information warfare environment
Authors: David P. Biros, Gregory Fields, Gregg Gunsch
Summary: Information use as a means of processing and exchanging information, increasing use, relying upon control, and operate as a critical function in society. This growing trend has generated sufficient interest in researchers to exam the behavior of people who rely on these information systems. This paper stated that most researchers deal primarily with interpersonal trust. However, with the advent of the computer age and with the increasing role that information systems play in society, there are a growing number of studies on the trust relationship between humans and information systems. Under the trust information systems positively influences trusting behavior study, the researchers use the Recognition-Primed Decision also known as (RPD). The model is used to describe the importance of the situation with respect to forming a decision. The literature also suggests that the more positive experiences decision-makers have with objects, like information systems, the more likely they will trust the object. In the empirical studies of the facet of trust, researchers proposes that the construct of external safeguards will have a positive effect on a decision-maker's trusting behavior
Source: http://csdl2.computer.org/comp/proceedings/hicss/2003/1874/02/187420062b.pdf

Title: Civilized machines: Constructing trust for artificial considerateness
Author: Chia-Hsun Jackie Lee
Summary: Traditionally, human and machines are living in different worlds. Both of them have their own cultures to communicate and interact with each other. Artificial Intelligence and Human-Computer interaction research fields have started blurring the boundary of human and machines' ecological realms. Phenomenon suggests that

perspective researchers can pay more attention to things that happened in action, rather than arguing what things are made of. According to Reeves and Nass people are easy to respond socially and naturally because those responses do not required strategic thinking and they are unconscious. In social constraint, humans expect machines should do every thing they can without violating social rules. Social constraints limit machines to do things good instead of doing the right thing. Social constraints represent the minimum social interaction in an abstract form that can be computed and stored, but social constraints cannot be made without humans. This paper investigates a theoretical framework for civilized machines. It also emphasizes how machine considerateness can be placed into human daily life. It also makes machine civilization shifting from proper politeness to socially acceptable considerateness. The paper suggests considerate acts that can proactively recognize situation and fulfill human needs.
Source: http://web.media.mit.edu/~jackylee/publication/hri_cmachine-lee.pdf

Title: How to trust robots further than we can throw them
Authors: David Bruemmer, Michael Goodrich, Nilanjan Sarkar, Bill Smart, Douglas Few, Donald Norman, Jean Scholtz, Mark L. Swinson, Holly Yanco
Summary: Humans have high expectations for mobile, intelligent robots. We have been told by film, fiction, and television that they will infiltrate our homes and places of work, changing the landscape of human experience. This panel investigates issues of trust surrounding implementation and use of mobile robots in a variety of different applications. The panel also showcases experiments and case studies that highlight the importance of operator trust. Panelist discusses how operator trust has affected the use and efficacy of their robots and show videos of their robots in action, to illustrate ways in which robots can fail. Some of the robots described represent a new frontier of research where robots and human work together as peers. Human-robot interaction is a complex and fascinating area for discussion. It leads to questions of a philosophical and ethical nature, and the panel also encourages this tendency. In summary the panelists have drawn somewhat different conclusions, depending on the subject of study. But in general the panelists compared the failure and success between human and machines interaction.
Source:
http://www.inl.gov/adaptiverobotics/autonomousbehaviors/pubs/bruemmer_chi2004.pdf

Title: Human-computer interaction for complex pattern recognition problems
Authors: Jie Zou, George Nagy
Summary: We review some applications of human-computer interaction that alleviative the complexity of visual recognition by partitioning it into human and machine tasks to exploit the differences between human and machine capabilities. Human involvement offers advantages, both in the design of automated pattern classification systems, and at the operational level of some image retrieval and classification tasks. Recent development of interactive systems has benefited from the convergence of computer vision and psychophysics in formulating visual tasks as computational processes. Computer-aided classifier design and exploratory data analysis are already well established in pattern recognition and machine learning, but interfaces and functionality are improving. On the operational side, earlier recognition systems made use of human talent only in reprocessing and in copying with rejects and current talent based image

interaction. In contrast, some visual object classification systems can exploit such interaction. They require, however, a domain specific visible model that makes sense to both human and computer.
Source: http://www.springerlink.com/content/t73n83567258q118/

Title: Semantic web interactions on internet relay chat
Authors: Jennifer Golbeck, Paul Mutton
Summary: Internet relay chat (IRC) is a chat system that has millions of users. IRC robots (bots) are programs that sit in chat rooms and provide different services to users. The IRC bots as a mechanism for human interaction with the Semantic web – specifically with web services and knowledge bases, is simple to program, has a intuitive, conversational interface for human users, and fits well with the inputs and outputs of Semantic web queries. This paper presents implementation of bots for interacting with web services and FOAF/Trust data models. We outline tools for implementation and present directions for future work to generalize the function of these bots.
Source: http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/9225/29243/01320122.pdf

Title: Integrating human-computer interaction development into SDLC: A methodology
Authors: Ping Zhang, Jane Carey, Dov Te'eni, Marilyn Tremaine
Summary: Incorporating a human-computer interaction (HCI) perspective into the systems development life cycle (SDLC) is critical to information systems (IS) success and in turn to the success of businesses. However, modern SDLC models are based more on organizational needs than human needs. The human-interaction aspects and information systems are considered far too little (only the screen interface) and far too late in the IS development process (only at the design stage). Thus, there is often a gap between satisfying organizational needs and supporting and enriching human users as they use the system for their tasks. This problem can be fixed by carefully integrating HCI development into the SDLC process to achieve a truly human-centered IS development approach. This tutorial presents a methodology for such human-centered IS development where human requirements for the whole system are emphasized. An example of applying such methodology is used to illustrate the important concepts and techniques.
Source:
http://melody.syr.edu/pzhang/publications/CAIS_05_Zhang_etal_HCI_SDLC.pdf

Title: A case for Shibboleth and grid security: Are we paranoid about identity?
Author: Mark Norman
Summary: The findings in this paper represent some of the output of the ESP-GRID project following the consultation of current Grid users regarding the future nature of grid computing. The project found that there was a clear purpose for Shibboleth in a future grid and for the majority of users; this would be a secure and improved experience of grid computing. Client-Based PKI remains suitable and desirable for power users and we must be careful of the means by which we mix these two access management technologies. PKI is currently used to define grid identities but these are problematically conflated with authorization. The Grid community should work harder to separate identity/authentication and authorization. This paper also questions whether we need

identity to be asserted throughout grid transactions in every use case. Currently, this is a solution to security the requirement: It should not be requirement in itself. We propose that the Grid community should examine methods for suspension of a rogue user's activities, even without identity explicitly stated to all parties. The project introduced the concept of a Customer-Service Provider model of grid use and has produced demonstrations at the University of Glasgow.
Source: http://www.allhands.org.uk/2006/proceedings/papers/638.pdf

Title: Trust in automation: Designing for appropriate reliance
Authors: John D. Lee, Katrina A. See
Summary: Automation is often problematic because people fail to relay upon it appropriately. Because people respond to technology socially, trust influences rely on automation. In particular, trust guides reliance, when complexity and unanticipated situations make a complete understanding of the automation impractical. This review considers trust from the organizational, sociological, interpersonal, psychological, and neurological perspectives. It considers how the context, automation characteristics, and cognitive processes affect the appropriateness of trust. The context in which the automation is used influences the automation performance and provides a goal-oriented perspective to assess automation characteristics along a dimension of attributional abstraction. These characteristics can influence trust through analytic, analogical, and affective processes. The challenges of extrapolating the concept of trust to people in trust in automation are discussed. A conceptual model integrates research regarding trust in automation and describes the dynamic of trust, the role of context, and the influence of characteristics. Actual or potential applications of this research include improved designs of system that require people to manage imperfect automation.
Source: http://www.engineering.uiowa.edu/~csl/publications/pdf/leesee04.pdf

Title: Trust and deception in mediation communication
Authors: Judee K. Burgoon, Gates Mathew Stoner, Joseph A. Bonito, Norah E. Dunbar
Summary: Guided by the international deception theory and the principle of interactivity, this investigation examined whether community modalities differentially affect the extent to which group members develop trust or are vulnerable to manipulation and deceit, based on the degree of interactivity the modalities afford. According to the principle of interactivity; involvement mutuality should increase as one moves from text-to- audio and audiovisual (AV) modalities, to face-to-face (FtF) communication. Under non-deceptive circumstances, greater interactivity should elicit corresponding increases in trust and credibility; under deceptive circumstances, it should produce greater truth biases and inaccurate detection of deceit. This effect should be partly mitigated in text and audio modality due to the presence of diagnostic deception indicators pairs were assigned to a truthful or deceptive condition in one of three mediated conditions, or in a face-to-face condition. In the deceptive condition, one member of each pair was enlisted to deceive during the interaction. Following discussion, participants rated their communicative behavior and the credibility of the truthful or deceptive actor. Truth bias and accuracy in judging deceptive information was calculated. Results are compared to previous findings from face-to-face deception. Implications for collaborative technologies are advanced.
Source: http://csdl2.computer.org/comp/proceedings/hicss/2003/1874/01/187410044a.pdf

Title: A new look at the dynamics of human-automation trust: Is trust in human comparable to trust in machines?
Authors: Poornima Madhavan and Douglas A. Wiegmann
Summary: The trust placed in automated diagnostic aids by the human operator is one of the most critical psychological factors that influence operator reliance on decision support systems. Studies examining the nature of human-interaction with automation have revealed that users have a propensity to apply norms of human-human interpersonal interaction to their interaction with "intelligent machine". Nevertheless, there exist subtle differences in the manner in which humans perceive and react to automated aids compared to human teammates. The present review is focused on comparing the process of trust development in human-automation teams with the human-human partnerships, specifically in the context of dyads that constitute a primary decision-maker and either a human "advisor" or an intelligent automated decision support system. A conceptual framework that synthesizes and contrasts the process of trust development in human verses automation is proposed. Potential implication of this research include the improved design of decision support systems by incorporating features into automated aids that elicit operator responses that mirror responses in human-human interpersonal interaction
Source:
http://www.humanfactors.uiuc.edu/Reports&PapersPDFs/humfac04/madwieg.pdf

Title: Assessment of operator trust in and utilization of automated decision-aids under different framing conditions
Authors: Ann M. Bisantz and Younho Seong
Summary: An analyst was used along with computerized software aids to explore the operators' abilities, inabilities and to detect the operators' failure rate in an industrial environment by using theoretical framework to describe potential factors affecting the operators' decision-making. The purpose of the experiment was to investigate the role of failure-cause on trust and system. Using two different sources of information, the information window and the decision-aid windows were analyzed and computed all the collected data. The results were used to provide some support for factors in the theoretical framework and demonstrate the use of an empirically developed trust scale.
Source:
http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V3143B3V1K-3&_user=10&_coverDate=08%2F31%2F2001&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=87b641e467357e8d80ef4137762d530e

Title: Effects of four computer-mediated communications channels on trust development
Authors: N. Bos, J.S. Olson, D. Gergle, G. Olson, Z. Wright
Summary: Studied the emergence of trust in a social dilemma game in four different communication situations: face-to-face, video, audio, and text chat. All three of the richer conditions showed significant improvements over text chat. Video and audio

conferencing groups were nearly as good as face-to-face, but both did show some evidence of what we term “delayed trust”.
Source:
http://www.soc.northwestern.edu/dgergle/pdf/BosOlsonGergleOlsonWright_RichMediaTrust_CHI02.pdf

Title: Attribution in distributed work groups
Author: Catherine Durnell Cramton
Source: In P.L. Hinds & S. Kiesler (Eds.), Distributed Work (pp. 191-212). The MIT Press.
Summary: The fundamental attribution error is likely to be exacerbated for dispersed collaborators relative to collocated collaborators, with grave consequences for group cohesion and viability. Working across dispersed locations typically reduces the situational information that collaborators have about each other, impacts how they process information, and fosters the development of in-groups and out-groups based on location. This biases perception of causes toward dispositional explanations, rather than situational explanations.

Title: The role of trust in automation reliance
Authors: M.T. Dzindolet, S.A. Peterson, R.A. Pomranky, L.G. Pierce, H.P. Beck
Summary: The study of trust and reliance, with the assistance of an automated decision-aid, participants viewed slides of Fort Sill terrain and indicated the presence or absence of a camouflaged soldier. The conclusion from the three studies indicates that trust is an important factor in understanding automation reliance decisions. Participants initially considered the automated decision-aid trustworthy and reliable. After observing the automated aid make errors, participants distrusted even reliable aids, unless an explanation was provided regarding why the aid might err. Knowing why, the aid might increase trust in the decision-aid and increased automation reliance, even when the trust was unwarranted. The studies suggest a need for future research focused on understanding automation use, examining individual differences in automation reliance, and developing valid and reliable self-report measures of trust in automation.
Source:
http://portal.acm.org/citation.cfm?id=941181&jmp=references&dl=portal&dl=ACM&CFID=11111111&CFTOKEN=2222222

Title: The equalization phenomenon: Status effects in computer-mediated and Face-to-Face decision-making groups
Authors: Durovsky, V.J., Kiesler, S., Sethna, B. (1991)
Summary: The study in a laboratory experiment, which compared face-to-face communication with electronic mail in decision-making groups whose members, differed in social status. We examined status in two ways: by varying the external status of group members and by varying the decision task to manipulate expertise. When the groups made decisions in face-to-face meetings, the high-status member dominated discussions over the three low-status members. Also, the high-status member more often was a "first advocate" in the face-to-face discussions, and first advocates were more influential than

later advocates. These status inequalities in face-to-face decision-making were pronounced just when the high-status member's expertise was relevant to the decision task. When the same groups made comparable decisions using electronic mail, status and expertise inequalities in participation were reduced. A striking and unexpected result was that "first" advocacy was shared by high- and low-status members in discussions using electronic mail. This behavior resulted in increased equality of influence across status and expertise. We discuss the implications of these results for research and for design of new communication technologies.
Source: http://www.leaonline.com/doi/abs/10.1207/s15327051hci0602_2

Title: Work group structures and computer support: A field experiment
Authors: J. D. Eveland and T. K. Bikson
Summary: The experiment created tasks for two different groups within a company, where the groups are equipped with two different types of technology. One group had full conventional office support; the other had addition networked microcomputers with electronic mail and routine office software. The experiment took place over a one year period. The experiment concluded that use of computer support for cooperative work results in both quantitative and qualitative changes; however effective participation in such electronically supported group requires significant investments of time and energy on the part of its members to master the technology and relatively high level of assistance during the learning process.
Source:
http://www.interactiondesign.org/references/conferences/proceedings_of_the_1988_acm_conference_on_computer-supported_cooperative_work.html

Title: Measurement of trust in complex and dynamic systems using a quantitative approach
Authors: Ananth Uggirala, Anand K. Gramopadhye, Brain J. Melloy, Joe E. Toler
Summary: In highly automated inspection systems today, humans act in a supervisory capacity, i.e., monitoring the process during normal operations and intervening to take manual control when necessary to override faulty automation. Research has concluded that the intervention behavior of supervisors is based on their trust in the automation. However, several problems remain with the current trust paradigm, the most important ones concerning the ambiguity and inaccuracy of the data. An original theoretical framework relating machine properties to operator perceptions is needed. To achieve this framework, machine properties need to be mapped to these perceptions through the development of quantifiable dimensions, referred to as uncertainty. Uncertainty, unlike trust, can be easily quantified, and therefore, can be related to system properties. This research attempts to measure uncertainty using a Line Length Experiment Module that simulates a real-life automated system monitored by an inspector. The experiment tests the ability of a human inspector to identify defects relative to the machine's ability to do so, rather than trying to gauge their competencies separately and subsequently comparing them. Three levels of uncertainty, established using National Institute of Standards and Technology (NIST) guidelines, were administered in the experiment, with the users rating their trust at each level through questionnaires. The results showed that the overall trust in the system had a significant inverse relationship with the system uncertainty.

Hence, the performance of hybrid systems can be improved by decreasing uncertainty, an improvement that will have an impact on quality.
Source:
http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V314CHJC0F1&_user=10&_coverDate=09%2F30%2F2004&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=74d9960c980fbe629f76a7cec64fbda7

Title: Trust a collision of paradigms
Authors: L. Jean Camp, Helenissen Baum, Cathleen McGrath
Summary: The technological challenges of securing networks are great, recently witnessed are widespread denial of service and virus attacks. Human reactions to these virus attacks may either be a loss of trust or willingness to tolerate increasing risk having experienced an assault. Examining human and computer-interaction with a focus on evaluations of human response to loss of trust, is a key part of the search for more secure networks. The success of current efforts to design appropriate security mechanisms depends as much on an understanding of human extensions of trust to computer as it does on an understanding of underlying mathematics. However, the former has not been sufficiently examined.
Source: http://itc.mit.edu/itel/docs/2001/trust_cnm.pdf

Title: Trust in close relationships
Authors: John K. Rempel, John G. Holmes, Mark P. Zanna
Source: Journal of Personality and Social Psychology, 1985, Vol. 49, No. 1, Pages 95-112
Environment: Sociological
Operational Measures: Evaluative Scales (trust and motivation)
Automation: No theories, trust is highly based on faith in one another and developed over time.
Experiments: Couples, cohabiting, newly-married and long-time married with children were the subjects of this experiment. They filled out questionnaires and materials independent of their partners and their answers were given based on scales.
Summary: Trust is "confidence that one will find what is desired from another, rather than what is feared" (Deutsch, 1973). Driscoll, Davis and Lipetz (1972) examined the link between trust and love and according to their thinking; trust evolves through mutually satisfying interactions and increasing confidence in the relationship. First, trust is seen to evolve out of past experience and prior interaction; thus it develops as the relationship matures. Second, dispositional attributions are made to the partner, such that he or she is regarded as reliable, dependable and concerned with providing expected rewards. Third, as Deutsch's use of the term fear implies, trust involves a willingness to put oneself at risk, be it through intimate disclosure, reliance on another's promises, sacrificing present rewards for future gains, and so on. These considerations point to a model of trust with three components that reflect increasing levels of attributional abstraction. We have labeled these components predictability, dependability and faith. The results showed that trust is related in important ways to the success of a close relationship. It is equally apparent that trust is a construct with a number of different elements and these elements do not all make equivalent contributions.

The most important aspect of trust in close relationships appears to be faith. In terms of both predictive and discriminant validity, faith had by far the strongest correlations with measures of love and happiness and was uniquely tied to perceptions of a partner's intrinsic motivation.

Title: Experimental comparisons of face-to-face and mediated communication: A review
Author: Ederyn Williams
Source: Psychological Bulletin, 1977, Vol. 8, No. 5. Pages 963-976
Environment: Distributed
Operational Measures: Analysis of Results by Evaluative Scales
Automation: No
Theories: Communication media replacing face-to-face communication
Experiments: Two-person teams were asked to solve problems by one of four communications media: (a) teletypewriting, (b) remote handwriting, (c) audio only, and (d) face-to-face. Major differences between the first two (written) and last two (voice) media were discovered. Problems took longer to solve in the two written modes, though fewer messages were exchanged and various differences in the activities (e.g., sending, receiving, writing) of the participants were identified. There were few differences between the two written modes and between the two voice modes. Similar results were obtained in other studies, using various other cooperative problems and different combinations of media. Another series of experiments studied conflicts of opinion. Subjects were asked about their opinion on social issues and then were paired. The paired members disagreed on at least two of the issues and were asked to discuss those issues until they reached an agreement by audio only, audio-video, or by face-to-face.
Summary: The study of human communication through electronic media is a relatively recent concern of psychologists. Telecommunications media, such as the telephone, closed circuit television, teleconference devices, and conferencing via computer terminals, are increasingly replacing face-to-face communication (Cherry, 1971; Day, 1973). A large body of research suggests that nonverbal cues have many indicative functions, in that they can act as cues to indicate the psychological state of the individual or the progress of an interaction (Argyle, 1969; Duncan, 1969; Mehrabian, 1972). However, researchers studying the communicative role of nonverbal cues (i.e., the extent to which they affect or alter the interaction processes or outcome) have encountered severe problems in manipulating these cues. Many studies (e.g., Ellsworth & Carlsmith, 1968) have used confederates with prespecified interaction patterns, such as continuous eye gaze, whereas others have adopted bizarre manipulations, such as having subjects wear cardboard masks or converse through a one-way mirror (e.g., Argyle, Lalljee, & Cook, 1968). Accordingly, some researchers have tuned to more naturalistic situations, which nevertheless alter the communication of nonverbal cues such as the comparison of familiar telecommunications media (e.g., telephone and face-to-face, by Cook & Lalljee, 1972). The range of media, from written communication, through the telephone and interactive television, up to face-to-face, allows the manipulation of all nonverbal cues, without the interaction appearing particularly bizarre to the participants. Contrary to the lack of media differences, when the tasks are cooperative or when the task is conflictful, several types of media differences appear. It was hypothesized that the more formal the medium of communication (audio only), the person with the stronger case would do better than he would do with a less formal medium (face-to-face). Also, it was found there

were significantly more total breakdowns in negotiations (i.e., failure to reach an agreement) with audio only than with face-to-face.

APPENDIX D

Trust Measures used in Jarvenpaa, Knoll, & Leidner (1998)

The items measuring benevolence, ability, integrity, trustworthiness, trust, and propensity to trust are as follows:

Benevolence
The other team members were very concerned about the ability of the teams to get along.
The outcomes of this project are very important to the other team members.
The other team members would not knowingly do anything to disrupt or slow down the project.
The other team members are concerned about what is important to the team.
The other team members will do everything within their capacity to help the team perform.

Integrity
The other team members try to be fair in dealing with one another.
The other team members have a strong sense of commitment.
I never am doubtful about whether the other team members will do what they promised.
I like the work values of the members on this team.
The other team members do not behave in a consistent manner – I am never sure if they are going to do what they promise or not.
The other team members display a solid with ethic.

Ability
I feel very confident about the other team members' skills.
The other team members have much knowledge about the work that needs to be done.
The other team members have specialized capabilities that can increase our performance.
The other team members are well qualified.
The other team members are very capable of performing their tasks.
The other team members seem to be successful in the activities they undertake.

Trust
If I had my way, I wouldn't let the other team members have any influence over issues that are important to the project.
I would be comfortable giving the other team members complete responsibility for the completion of this project.
I really wish I had a good way to oversee the work of the other team members on the project.
I would be comfortable giving the other team members a task or problem that was critical to the project, even if I could not monitor them.

Trustworthiness
Members of my work group show a great deal of integrity.
I can rely on those with whom I work in this group.
Overall, the people in my group are very trustworthy.
We are usually considerate of one another's feelings in this work group.

The people in my group are friendly.
There is no “team spirit” in my group.
There is a noticeable lack of confidence among those with whom I work.
We have confidence in one another in this group.

Propensity to Trust
One should be very cautious when working with foreign students.
Most foreign students tell the truth about the limits of their knowledge.
Most foreign students can be counted on to do what they say they will do.
If possible, it is best to avoid working with foreign students on projects.
Most foreign students are honest in describing their experience and abilities.
Most foreign students answer personal questions honestly.
Most foreign students are very competent in terms of their studies.

APPENDIX E

Summary Table from Powell, Piccoli, & Ives (2004)

Authors	Year	Issues/variables examined	Results	Theory	Time frame and subjects	Team size	Journal
Ahuja & Carley	99	Task, network structure, hierarchy, centralization, fit, performance	VTs can be hierarchical and centralized from a communication standpoint. Fit between task and structure not associated with objective performance, but is associated with satisfaction	Network & organization form theory	Ongoing Organization employees and academics	66	Organization Science
Archer	90	Decision quality, # alternatives, tome to reach decision, satisfaction	VT took longer to reach decision, bit had equal quality and satisfaction	Results provided support for AST	8-weeks grad students	4-5	Behavior & Information Technology
Berdahl & Craig	96	Participation, influence, gender	Participation more centralized in VTs. In VTs, males in majority–female teams had more influence, males in majority–male teams had less influence than females	Proportional, Social Role, Expectation states	7-weeks UG students	3-4	Journal of Computer Supported Collaborative Work
Burke & Aytes	98	Cohesiveness, equality of participation, leadership, coordination, performance	No difference between VT and TT (traditional teams) in cohesiveness, performance and equality of participation. Leadership better in TT. VT better at coordination at first		4-weeks UG students	3-4	HICCS

Burke & Chidambaram	96	Social presence, communication effectiveness, ease-of-use, performance	TT reported greater social presence, communication effectiveness, and ease-of-use. No difference in performance between VT and TT	Social Presence & Media Richness	4-weeks UG students	5	HICCS
Chidambaram	96	Work processes, cohesiveness, satisfaction	Over time, VT members will exchange enough information to develop relational ties. Over time, attitudes improve as does satisfaction with outcomes.	SIP, Punctuated Equilibrium Model	4 week UG students	5	MIS Quarterly
Chidambaram & Bostrom	93	Conflict management, cohesiveness, alternatives, decision quality	No difference between teams' decision quality, VT more equal participation, task focus, more alternatives generated. TT better at first in conflict mgmt, cohesiveness; by midpoint, VT had surpassed TT	SIP, AST	4-weeks UG students	5	Journal of Organization Computing
Chidambaram et al.	90-91	Conflict resolution, cohesiveness	Conflict resolution and cohesiveness both initially better in TT, by time 4, VT better at both VT more likely to think of idea as "ours"	AST	4-weeks UG students	5	JMIS
Crampton	01	Mutual knowledge, communication problems	5 types of communication problems identified from failure of creating mutual knowledge un VT	Communication Theories	7-weeks Grad students	6	Organization Science
Eveland & Bikson	88	Leadership, structure, satisfaction,	TT greater leadership stability. VT leadership fluctuates more. VT created		1-year Org workers	20 6-10 in sub-	ACM Transactions on OIS

		communication	unique structure. VT satisfaction continually increased. TT satisfaction remained same. VT communicated more.		& retirees	groups	
Galegher & Kraut	94	Planning, social conversations, revising/writing, coordination difficulty, total communication, communication quality, perceived fairness, perceived project quality, project performance	TT more effective planning, revising/writing, perceived fairness, communication quality. VT had more total communication. No difference in project performance. VT required more time, spent more time communicating. VT had more coordination difficulty, fewer social conversations	Contingency Theory Discuss several others	4-weeks Grad students	3	Information Systems Research
Galvin & Ahuja	01	Communication, member status, info exchange, info content	Newcomers to team engage in greater info seeking, established members engage in more info providing. Established members engage more in seeking norms & values info – new members more likely to seek out this type of info from “private” source		Ongoing Org employees & academics	Not given	Our Virtual World chapter
Hollingshead, McGrath, & O’Connor	93	Task type, task performance, cumulative experience, changes	No difference between VT and TT for generating and decision-making tasks. TT better for negotiation and intellective tasks early on. Relationship between task	Task–media fit Task circumplex model	9-weeks UG students	4-5	Small Group Research

			and technology performance more dependent on experience with technology and team membership rather than task type				
Jarvenpaa, et al.	98	Benevolence, integrity, ability, team building, trust, propensity to trust others	Integrity early on important to development of trust, benevolence over time. Team building increased knowledge about team members, but did not directly increase trust	Results support theory of swift trust	8-weeks Grad students	4-5	JMIS
Jarvenpaa & Leidner	99	Swift trust	VTs with high trust also had: social communication, enthusiasm, predictable communication, substantial & timely feedback, ability to cope with tech uncertainty, initiative, social to task focus, positive leadership, phlegmatic response to crises. Swift trust likely a result of communication	Swift trust	6-weeks Grad students	4-5	Organization Science
Johansson et al.	99	Communication, coordination	Communication and coordination were biggest issues in VT. Obstacles to coordination: power, doing it own way (power issues), communication predictability, culture		5-weeks UG and Grad students	22	IEEE Transactions on Professional Communication

Kaiser et al.	00	Performance	To improve performance in VT: intermediate goals as well as final goals, similar training for all, clear structure, team maintenance activities/team building		3-weeks UG and Grad students	6	Business Communication Quarterly
Kayworth & Leidner	00	Communication, culture, technology, leadership, relational links, satisfaction, success	VTs using more communication methods more satisfied and more successful. Cultural differences intensify communication and coordination problems. Effective leadership needed, richer communication facilitates socialization		6-weeks UG and Grad students	5-7	European Mgmt Journal
Kayworth & Leidner	01-02	Leader role, role clarity, leadership effectiveness, culture, communication effectiveness	Effective leadership associated with communication effectiveness, communication satisfaction, role clarity, better relational skills, ability to foster relational aspects, structure	Leadership Theories	5-weeks UG and Grad students	5-7	JMIS
Krumpel	00	Group knowledge production, communication content	Group knowledge production possible in VT. Group engaged in knowledge production would do better with a leader	Structuration Theory	Ongoing Working Group of subcommittee ‘e’ for IEEE	51	IEEE Transactions on Professional Communication
Lind	99	Gender. Cohesion, conflict, quality of work, inclusiveness	Women more satisfied, felt more included and greater team cohesiveness with the		5-6 weeks Students	4	IEEE Transaction on Professional

			VT experience than men. No difference in quality of work between VT and TT				Communication
Lurey & Raisinghani	01	Job characteristics, selection procedures, team relations, team processes, leadership, effectiveness, education, rewards, executive leadership, tools & technologies, communication	Team processes and member relations had strongest relationship to performance and satisfaction. Selection procedures and executive leadership style moderately related to performance and satisfaction		Length varied Organization employees	Varied (3-15 in 12 separate teams)	Information and Management
Majchrzak et al.	00a	Technology use, success, task type, conflict, shared language/knowledge	FtF or phone used for ambiguous tasks, managing conflicts, brainstorming, clarifying goals, etc. VT used for routine tasks of analysis, project status. FtF meetings early on created a shared language between members – this enabled ambiguous tasks to be completed later by VT	Media Richness, Social Presence, Task Circumplex	10-months Organization employees	8	Information Resources Mgmt Journal
Majchrzak et al.	00b	Adaptation, technology, structure, knowledge sharing, decision-making	Using collaborative technology creates several misalignments with pre-existing structure and environment. VTs need to adapt all 3 structures – technology, org environment,	AST	10-months Organization employees	8	MIS Quarterly

			and work group structures				
Malhotra et al.	00	Shared understanding, knowledge sharing	Three mgmt practices contribute to success of VT: strategy-setting, technology use should facilitate shared knowledge and collaborative use, restructuring work without changing core needs		10-months Organization employees	8	MIS Quarterly
Mark	01	Socialization, technology, participation, team culture, integrating team members	Key challenges of VT of participation, team culture and integrating remote work discussed		Over 6-months Organization employees	8-15	Our Virtual World chapter
Maznevski & Chudoba	01	Processes, performance, communication, cohesion, culture, technology use, decision quality, commitment	Effective VTs have high commitment, high cohesiveness, effective processes, satisfaction with decisions, continuous communication, and task-technology fit. Cultural differences affected coordination. Regular FtF meetings provided necessary coordination	AST	Long-term Organization employees	8, 9, & 12	Organization Science
McDonough et al.	01	Behavioral challenges, project mgmt challenges, performance	Behavioral challenges greatest in GVT, then VT, then TT. Project mgmt challenges least in TT. Performance greatest in TT, then VT, then GVT. Greater project mgmt challenges are associated with lower		Ongoing Organization employees	Did not report	Journal of Product Innovation Management

			performance for all 3 types of teams. Project mgmt challenges more a function of distance between members than cultural differences				
Montoya-Weiss et al.	01	Process structure, conflict, performance	Avoidance, compromise conflicts have negative relationship with performance. Process structure weakens negative effect on performance for avoidance, compromise conflict. Competition, collaboration conflicts have positive relationship with performance	Conflict Theories	3-weeks UG students	5	Academy of Management Journal
Qureshi & Vogel	01	Adaptation (technology, work, social), structure, specialization, coordination, task, learning	Provide framework of adaptation and organization challenges in VTs. Review of current research on adaptation	Structuration Theory	Not applicable	NA	Group Decision and Negotiation
Ramesh & Dennis	02	Coordination, work processes	VTs may operate better as object-oriented teams – decoupling team members through use of well defined processes, inputs, and/or outputs rather than trying to tightly couple members may improve coordination in VTs	Media Richness, Media Synchronicity	3-months, 11-months, 2-years Organization employees	3 6 10-58	HICCS

Robey et al.	00	Learning, mgmt, performance, communication, technology use, FtF meetings	Communication improves cultural understanding which improves coordination. FtF meetings needed to gain respect, trust, and interpersonal relations. Choice of technology based on many factors. Learning can be effectively accomplished in VT.	Learning Theory	3+ years Organization employees	11-13	IEEE Transactions on Professional Communication
Sarker & Sahay	02	Time, space, culture, technical expertise, communication, coordination, socialization	Strategies for dealing with challenges involved with VTs		14-weeks Students	8-10	HICCS
Sarker et al.	01	Technology, norms, social practices, team development stages, frames of reference, collaboration	Communication, collaboration, social norms all adapt and change in different stages of team development	Adapted Grounded Theory Communication Theories Structuration	14-weeks Students	8-10	Database
Savivki et al.	96	Gender, task, communication, team development	Female-only VTs are more satisfied, send more words, have better team development, and self-disclose more than male-only or mixed-gender VTs.		4-weeks UG students	4-6	Computers in Human Behavior
Sharda et al.	88	Effectiveness, decision time, confidence, #	VT greater effectiveness and took longer to make decision. No difference between VT		8-weeks UG students	3	Management Science

		alternatives	and TT on confidence, # alternatives generated				
Suchan & Hayzak	01	Communication, attitudes, technology use	Communication most important factor for success, seen as strategic activity to be considered daily. Mentoring program enabled socialization of members. Leaders used FtF meetings to help develop trust. Culture and reward system supported information-sharing		Ongoing Organization employees	31	IEEE Transactions on Professional Communication
Tan et al.	00	Cohesiveness, collaboration, perceived decision quality, perceived decision satisfaction	VTs receiving dialogue training had higher cohesiveness, collaboration, perceived decision quality, perceived decision satisfaction. Differences remained over time	Dialogue Theory	4-weeks UG students	5	IEEE Transactions on Professional Communication
Van Rysson & Godar	00	Culture, technology, communication effectiveness	Cultural differences created problems for UG students in areas of socialization, communication, coordination. Training needed both in how to communicate and how to use technology		3-months UG students	4	Journal of International Management
Walther	95	Relational communication	VTs became less task-oriented and less formal over time. VTs perceived greater immediacy/affection/	SIP	5 week UG students	3	Organizational Science

			inclusion, communication of attitude likeness than did TT				
Walther & Burgoon	92	Relational communication	VTs increased in several relationship dimensions to more positive levels – approaching or surpassing TT levels	SIP SPT	5-weeks UG students	3	Relational Communication
Warkentin & Beranek	99	Communication training, interaction process, trust, commitment to team goals	By mid-point of VT life, teams with communication training had higher trust, commitment to team goals, openness of expression, better interaction processes	Discuss many	6-weeks Grad students	3-4	Information System Journal
Warkentin et al.	97	Performance, relational links, info exchange	TT had higher relational links, performance. No differences on effectiveness of info exchange	TIP	3-weeks UG students	3	Decision Science

APPENDIX F

Summary of Major Issues Investigated from Powell, Piccoli, & Ives (2004)

			Inputs				**Socio-emotional Processes**			**Task Processes**			**Outputs**	
Authors	Year	Subject	Structure	Culture	Technology	Training	Cohesiveness	Trust	Relationship Building	Communication	Coordination	Task-Technology-Structure Fit	Performance	Satisfaction
Ahuja & Carley	99	O	X							X			X	
Archer	90	S											X	X
Berdahl & Craig	96	S								X				
Burke & Aytes	98	S					X			X	X		X	
Burke & Chidambaram	96	S			X				X	X			X	
Chidambaram	96	S					X				X	X		X
Chidambaram & Bostrom	93	S					X		X				X	
Chidambaram et al.	90-91	S			X		X					X		
Crampton	01	S								X				
Eveland & Bikson	88	O	X							X				X

Galegher & Kraut	94	S	X						X	X	X		X	
Galvin & Ahuja	01	O	X							X				
Hollingshead, McGrath, & O'Connor	93	S			X							X	X	
Jarvenpaa, et al.	98	S			X			X	X	X				
Jarvenpaa & Leidner	99	S	X					X						
Johansson et al.	99	S		X					X	X	X			
Kaiser et al.	00	S	X			X							X	
Kayworth & Leidner	00	S	X	X	X				X	X	X	X	X	X
Kayworth & Leidner	01-02	S	X	X					X	X				
Krumpel	00	O	X											
Lind	99	S					X		X		X		X	X
Lurey & Raisinghani	01	O							X	X	X	X		
Majchrzak et al.	00a	O	X									X	X	
Majchrzak et al.	00b	O	X								X	X		
Malhotra et al.	00	O	X								X		X	
Mark	01	O		X	X				X	X				
Maznevski & Chudoba	01	O	X	X			X			X	X	X	X	

McDonough et al.	01	O		X				X	X	X			X	
Montoya-Weiss et al.	01	S	X										X	
Qureshi & Vogel	01	N/A	X							X	X	X	X	
Ramesh & Dennis	02	O	X							X	X	X		
Robey et al.	00	O	X	X					X	X	X	X	X	
Sarker & Sahay	02	S		X		X			X	X	X			
Sarker et al.	01	S	X		X			X	X		X	X		
Savicki et al.	96	S							X	X				X
Sharda et al.	88	S										X	X	
Suchan & Hayzak	01	O	X	X		X		X		X			X	
Tan et al.	00	S				X	X				X		X	X
Van Rysson & Godar	00	S		X	X	X				X			X	
Walther	95	S							X					
Walther & Burgoon	92	S						X	X					
Warkentin & Beranek	99	S				X		X			X			
Warkentin et al.	97	S					X		X	X	X		X	X